Investigating Protein Conformational Change Via Molecular Dynamics Simulation

A thesis submitted to the University of Manchester for the degree of PhD in the Faculty of Medical and Human Sciences

2011

Neil John Bruce

School of Pharmacy and Pharmaceutical Sciences

Contents

List of Figures

List of Tables

Abstract

Submitted by **Neil John Bruce** for the Degree of PhD, entitled *"Investigating Protein Conformational Change Via Molecular Dynamics Simulation"*. 10 May 2011.

Accumulation and aggregation of the 42-residue amyloid-β (Aβ) protein fragment, which originates from the cleavage of amyloid precursor protein by β and γ secretase, correlates with the pathology of Alzheimer's disease (AD). Possible therapies for AD include peptides based on the Aβ sequence, and recently identified small molecular weight compounds designed to mimic these, that interfere with the aggregation of Aβ and prevent its toxic effects on neuronal cells in culture. Here, we use molecular dynamics simulations to compare the mode of interaction of an active (LPFFD) and inactive (LHFFD) β-sheet breaker peptide with an Aβ fibril structure from solid state NMR studies. We found that LHFFD had a weaker interaction with the fibril than the active peptide, LPFFD, from geometric and energetic considerations, as estimated by the MM/PBSA approach. Cluster analysis and computational alanine scanning identified important ligand-fibril contacts, including a possible difference in the effect of histidine on ligand-fibril π-stacking interactions, and the role of the proline residue establishing contacts that compete with those essential for maintenance of the inter-monomer β-sheet structure of the fibril. Our results show that molecular dynamics simulations can be a useful way to classify the stability of docking sites. These mechanistic insights into the ability of LPFFD to reverse aggregation of toxic Aβ will guide the redesign of lead compounds, and aid in developing realistic therapies for AD and other diseases of protein aggregation. We have also performed long explicit solvent MD simulations of unliganded amyloid fibril in three putative protonation states, in order to better understand the energetic and mechanical features of the fibril receptor. Over 100 ns MD simulations, the trajectories where fibril has Glu11 and Glu22 side-chains protonated exhibit the least deviation from the initial solid state NMR structures. Free energy calculations on these trajectories suggest that the weakest fibril interface lies in the lateral rather than transverse direction and that there is little dependence on whether the lateral interface is situated at the edge or middle of the fibril. This agrees with recent reported steered molecular dynamics calculations.

Secondly, in an effort to improve the ability of atomistic simulation techniques to directly resolve protein tertiary structure from primary amino acid sequence, we explore the use of a molecular dynamics technique based on swarm intelligence, called SWARM-MD, to identify the native states of two peptides, polyalanine and AEK17, as well as Trp-cage miniprotein. We find that the presence of cooperative swarm interactions significantly enhanced the efficiency of molecular dynamics simulations in predicting native conformation. However, it also is evident that the presence of outlying simulation replicas can adversely impact correctly folded replica structures. By slowly removing the swarm potential after folding simulations, the negative effect of the swarm potential can be alleviated and better agreement with experiment obtained.

Declaration

The author declares that no portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

i The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the "Copyright") and he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made **only** in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the thesis, for example graphs and tables ("Reproductions", which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see `http://www.campus.manchester.ac.uk/medialibrary/policies/intellectual-property.pdf`), in any relevant Thesis restriction declarations deposited in the University Library, The University Library's regulations (see `http://www.manchester.ac.uk/library/aboutus/regulations`) and in The University's policy on presentation of Theses.

The Author

The author graduated with an MChem in Chemistry from the University of Sheffield in 2005. In 2007 he graduated with an MSc in molecular modelling from Cardiff University. Since then he has been undertaking the work outlined in this thesis. He is the first-author of two peer reviewed papers,[1;2] which form the basis of sections 3.2 and 4.3 of this work.

Acknowledgements

The author would like to express his gratitude to number of people, without whom completing this thesis would not have been possible.

He would especially like to thank his supervisor Dr Richard Bryce for his continued advice and guidance, while performing the work outlined in this thesis, and while preparing the thesis.

He would also like to thank all the people who have been involved with the Computational Biophysics and Drug Design group, past and present, both for the support they have given, and for making my time in the lab enjoyable.

He would like to thank Research Computing Services at the University of Manchester, the National Grid Service and the National Service of Computational Chemistry Software for providing the computational resources that were required for this work. He would particularly like to thank Pen Richardson at Research Computing Services for the support she has given.

He is grateful for the financial support that has been provided by the Engineering and Physical Sciences Research Council.

Finally, he would like to thank his family and friends for their love and support.

Chapter 1

Introduction

Biomolecules such as proteins and peptides can be highly flexible. Their flexibility can have important consequences on their function as well as the ability to target them via computer-aided design. The aim of this thesis is to use computational simulation methods in order to explore the structure and stability of biomolecules and their complexes. In the following Chapter, we outline the theoretical background to the methods employed in this work. In Chapter 3, we introduce Alzheimer's disease and study the targeting of Aβ amyloid fibrils, implicated in the disease. Using molecular dynamics, we examine the binding of active and inactive peptides to an amyloid fibril model. We also use simulations and free energy calculations in order to deepen our understanding of the stability of amyloid fibril in the absence of ligand. In Chapter 4, we examine enhanced sampling molecular dynamics approaches and their application to biomolecular conformation. Here, we implement and apply the SWARM-MD method, a molecular dynamics approach based on simulating system replicas in parallel, in order to identify the native state of biomolecules, using as examples two peptides and a mini-protein.

Chapter 2

Theory and Methods

Over the last few decades, the power and availability of computational resources within the scientific community has increased greatly. This has led to the development and exploitation of many computational chemistry techniques, which can be used both to explain experimental results and guide future research. In this chapter, the theoretical background to the methods employed in this work will be discussed.

2.1 Molecular Modelling

2.1.1 Quantum Mechanics Background

Computational chemistry techniques seek to create a theoretical framework within which the properties of atomic and molecular systems can be described. Of fundamental importance to this is the calculation of how the energy of the system can be obtained from its current state. The energy of a system of particles is given by the Hamiltonian H:

$$H(\mathbf{p}, \mathbf{r}) = T(\mathbf{p}) + V(\mathbf{r}) \tag{2.1}$$

where T is the total kinetic energy of the particles in the system as a function of their momenta $\mathbf{p}$, and V is the total potential energy of the particles as a function of their positions $\mathbf{r}$.

In a chemical problem, the particles of interest are the electrons and nuclei of the chemical species involved. At the scales involved in these interactions, it is necessary to invoke a quantum mechanical description of the system. In the quantum Hamiltonian $\hat{H}$, the kinetic and potential energy functions are replaced

by mathematical operators, $\hat{T}$ and $\hat{V}$:

$$\hat{H} = \hat{T} + \hat{V} = -\frac{\hbar^2}{2m}\nabla^2 + \hat{V} \tag{2.2}$$

which act on the wavefunction Ψ of the system to describe its evolution via the Schrödinger equation:

$$i\hbar\frac{\partial}{\partial t}\Psi = \hat{H}\Psi \tag{2.3}$$

The energy E of a stationary state in a quantum system is obtained from the eigenvalues of the time-independent Schrödinger equation:

$$E\Psi = \hat{H}\Psi \tag{2.4}$$

Quantum calculations are rarely performed using the full molecular Hamiltonian, which depends on the positions and momenta of all electrons and nuclei in the system. A realisation that the masses of the nuclei are orders of magnitude greater than that of an electron led to the Born-Oppenheimer approximation. This states that the electrons are able to rearrange themselves instantaneously following rearrangements in nuclear positions. This allows the nuclear and electronic wavefunctions to be separated, and electronic calculations to be performed assuming that the nuclei are fixed charges in space. Solving the electronic time-independent Schrödinger equation for different nuclear positions leads to the generation of a potential energy surface, which describes how the potential energy of the system changes when varying atomic coordinates.

2.1.2 Parameterised Force fields

Performing a quantum mechanical calculation, even when operating under the Born-Oppenheimer approximation, is time consuming. This is due to both the need to include the interactions between all electrons and nuclei in the system, and the complex numerical techniques required to solve the time-independent Schrödinger equation. Force field–based molecular mechanics techniques seek to reduce the computational cost of these calculations by reducing the dimensionality of the problem, as only nuclear coordinates are required, and significantly reducing the complexity of the functions required in the calculation.

A molecular force field consists of a series of classical potentials that describe how the potential energy of the system changes with respect to changes in internal molecular coordinates, relative to reference equilibrium positions. Taken as a whole, the sum of these potentials should approximate the quantum potential energy for the system of interest, at least where there are no significant deviations from the reference positions used in the potential terms. As electronic effects are excluded from these calculations, they are only applicable to situations were there would be no change in electronic structure. Thus they are unable to describe changes in the chemical bonding of the system.

The functional form of these potentials varies amongst different force fields, but the general scheme followed by most force fields is given by:

$$E = \sum_{\text{bonds}} E_{bond} + \sum_{\text{angles}} E_{angle} + \sum_{\text{dihedrals}} E_{dihedral} + \sum_{\text{atom pairs}} [E_{Coulomb} + E_{vdW}] \tag{2.5}$$

The first three terms in this equation are known as the bonded potentials, as they

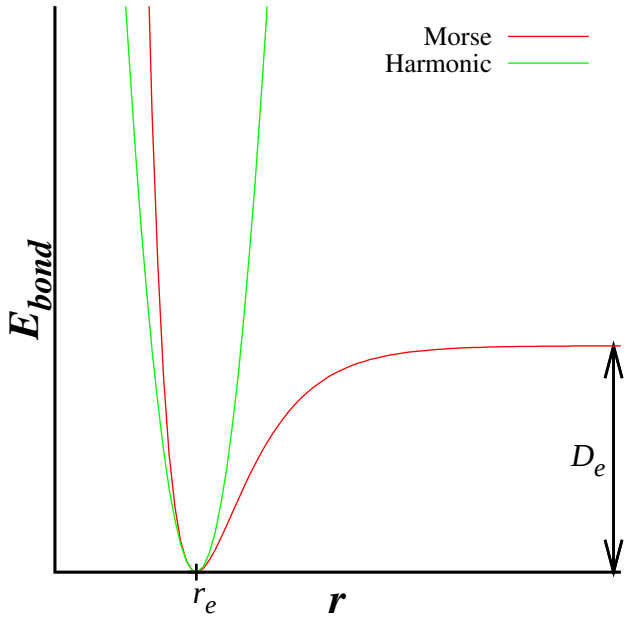


Figure 2.1: Morse and harmonic functions used to describe bond stretch potential energy curves

apply to atoms that are directly connected by chemical bonds; while the latter two are non-bonded potentials that apply to all pairs of atoms in the system.

The dependence of potential energy on the separation r of two chemically bonded atoms (Figure 2.1) is well described by the Morse potential:

$$E_{bond}(r) = D_e \left(1 - e^{-a(r-r_e)}\right)^2 \tag{2.6}$$

where D_e is the bond dissociation energy, r_e is the equilibrium bond length and a is a parameter that describes how quickly the potential changes with increasing separation.

Despite the accuracy of the Morse potential, bond stretch terms (E_{bond}) are most

usually represented by a harmonic function analogous to Hooke's law:

$$E_{bond}(r) = \frac{k}{2}(r - r_e)^2 \tag{2.7}$$

where k is the Hooke's law force constant. Although a poorer representation of the actual potential energy curve, this function is usually preferred due to the removal of the computationally expensive exponential function and its requirement of only two parameters, as opposed to the three required by the Morse potential. In the region close to the reference bond length r_e, the harmonic function gives a reasonable representation of the bond stretch potential energy curve.

Equation 2.7 is obtained from a Taylor expansion in r about the point r_e, truncated at the quadratic term (the first term in the expansion is a constant, which only serves to shift the potential to higher or lower energy, so can be ignored; while the second term is zero as the derivative of E_{bond} is zero at its minimum r_e). Some force fields improve the accuracy of the E_{bond} potentials by including higher order terms in the expansion, allowing bond anharmonicity to be included. The addition of just a cubic term leads to the bond energy tending towards negative infinity at large separations, but this error can be prevented by the addition of a quartic term. The resulting equation takes the form:

$$E_{bond}(r) = \frac{k}{2}(r - r_e)^2 \left(1 - k'(r - r_e) - k''(r - r_e)^2\right) \tag{2.8}$$

where k' and k'' are the anharmonicity constants of the cubic and quartic terms.

The angle bending term (E_{angle}) that describes how the potential energy varies with respect to the angle θ between three consecutively bonded atoms is also

generally modelled as a harmonic function:

$$E_{angle}(\theta) = \frac{k}{2}(\theta - \theta_e)^2 \tag{2.9}$$

As with the E_{bond} term, a better agreement with the true potential energy surface may be obtained by using a series of terms in higher orders of θ.

The dihedral angle term ($E_{dihedral}$) describes the change in potential energy on rotating a chemical bond around a torsional angle ϕ defined by four consecutively bonded atoms. A common form for this potential is:

$$E_{dihedral} = \frac{V}{2}(1 + \cos(n\phi - \gamma)) \tag{2.10}$$

where V gives the height of the barrier to rotation, n gives the periodicity and γ sets the position of the potential minima. For most molecular dihedral angles, different values are required for the energy of each minimum and barrier. For this reason, each dihedral is usually represented as a sum of terms like that in Equation 2.10. In addition to dihedral terms for four consecutively bonded atoms, improper torsional angle terms may also be present in the force field, to describe the energy penalty caused by out-of-plane bending, such as that by atoms bonded to aromatic rings. Improper torsions may either be treated with dihedral angle terms like Equation 2.10, or by a harmonic potential that depends on the angle between the plane of three atoms and a fourth atom.

The electrostatic interaction between a pair of charged atoms a and b ($E_{Coulomb}$) is described by the Coulomb potential:

$$E_{Coulomb} = \frac{q_a q_b}{4\pi\varepsilon_0 r_{ab}} \tag{2.11}$$

where r_{ab} is the separation between atoms a and b, and q_a and q_b are their partial atomic charges. The partial atomic charges are empirical parameters that must be fitted to match either experimental or quantum mechanical data.

The van der Waals term (E_{vdW}) models the attractive dispersion and Pauli repulsion interactions between the clouds of electron density on a pair of atoms, separated by a distance r. This interaction is most commonly modelled using a 6-12 Lennard-Jones potential.[3;4] One form of the 6-12 Lennard Jones potential is given by:

$$E_{vdW} = \varepsilon \left[\left(\frac{r_0}{r} \right)^{12} - 2 \left(\frac{r_0}{r} \right)^{6} \right] \tag{2.12}$$

where r_0 is the atomic separation at the minimum of the potential, and ε is the well depth of the potential. While the use of an r^{-6} dependent dispersion term is theoretically justified,[5] there is no theoretical justification for the use of an r^{-12} dependent repulsion term, which tends to overestimate the repulsive force at shorter distances. Instead, the r^{-12} term is often used as it is computationally efficient, being the square of the r^{-6} term used for the dispersion interaction. Alternatively, the Pauli repulsion term can be modelled using an exponential term, leading to the Buckingham potential:[6]

$$E_{vdW} = Ae^{Br} - \frac{C}{r^6} \tag{2.13}$$

While the Buckingham potential gives a more accurate representation of the repulsive energy wall close to the potential energy minimum, at small separations the potential passes through a maximum value, then tends toward negative infinity as the separation decreases (Figure 2.2). This leads to the possibility of atoms fusing if they come too close together. For this reason, and due to the presence of the computationally expensive exponential term, a Lennard Jones potential is

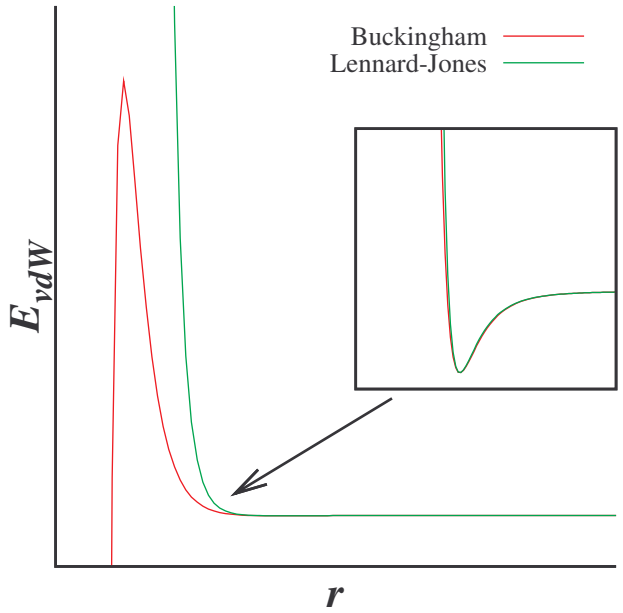


Figure 2.2: Comparison of Lennard-Jones and Buckingham potentials at short range. The Lennard-Jones potential rises too rapidly with decreasing separation; while the Buckingham potential passes through a maximum.

employed in most force fields.

The non-bonded force field terms $E_{Coulomb}$ and E_{vdW} are generally ignored for atom pairs that are directly connected by a covalent bond, as at short separations these potentials may give erroneous results, particularly when a r^{-12} dependent potential is used to model Pauli repulsion, and interactions between these atoms are already accounted for by the E_{bond} terms. Interactions between atoms separated by two covalent bonds (1,3–interactions) are also ignored as they are accounted for by the E_{angle} term. Similarly, the non-bonded interactions of atoms that are connected by three covalent bonds, so called 1,4–interactions, are often scaled down, as these interactions are already accounted for by the $E_{dihedral}$ term. Alternatively, some force fields may not include an $E_{dihedral}$ term, with the barriers to bond rotation modelled purely by the non-bonded terms.

2.1.3 Treatment of Solvent

When modelling biologically relevant systems, we are usually concerned with the properties of molecules in the condensed phase, most commonly in aqueous solutions. It is therefore necessary to include the effects that the solvent environment have on the molecules in the system. The methods by which we can model solvent interactions can be separated into two broad categories; explicit modelling of the solvent molecules that surround the solute, or implicit modelling of the solvent by surrounding the solute in a dielectric medium that represents the solvent environment.

2.1.3.1 Explicit Solvent Models

When modelling explicit water, the number of interactions present in the system can increase significantly. For this reason it is preferable to have as simple a representation of the water molecules as possible. Two of the most commonly used water models are the TIP3P[7] and SPC[8] models. In both of these models, water is represented by positive charges on the hydrogen atoms and a negative charge on the oxygen. There is only one set of van der Waals parameters per model, with the interaction site centred on the oxygen atom. The O–H bonds of water are kept rigid. SPC/E[9] is an updated version of the SPC model.

2.1.3.2 Implicit Solvent Models

Using an implicit representation of the aqueous environment can significantly reduce the dimensionality of the system, by reducing the number of non-bonded interactions that need to be calculated. There are a number of solvent-solute

interactions that need to be accounted for in the implicit model. The free energy of solvation of a solute can be given by:

$$\Delta G_{solv} = \Delta G_{elec} + \Delta G_{disp} + \Delta G_{rep} + \Delta G_{cav} \tag{2.14}$$

where ΔG_{elec} is the free energy change caused by the electrostatic interactions between the solvent and the solute; ΔG_{disp} and ΔG_{rep} are the free energy changes due to the dispersion and repulsion van der Waals interactions between the solvent and the solute; and ΔG_{cav} is the cavitation energy penalty caused by the need to create a cavity in the solvent to house the solute.

ΔG_{elec} can be found by solving the Poisson equation, or for an aqueous solution containing mobile ions, the Poisson-Boltzmann equation:[10]

$$\nabla \cdot \varepsilon(\mathbf{r}) \nabla \phi(\mathbf{r}) - \kappa' \sinh[\phi(\mathbf{r})] = -4\pi\rho(\mathbf{r}) \tag{2.15}$$

where ε is the dielectric constant of the solvent, ϕ is the electrostatic potential and ρ is the charge density. κ' is related to the Debye-Hückel inverse length κ, which relates to the ionic strength of the solution. The Poisson-Boltzmann equation can be linearised by performing a Taylor expansion of the hyperbolic sine function, and truncating this expansion after the first term:

$$\nabla \cdot \varepsilon(\mathbf{r}) \nabla \phi(\mathbf{r}) - \kappa' \phi(\mathbf{r}) = -4\pi\rho(\mathbf{r}) \tag{2.16}$$

This equation is commonly solved used the finite-difference–Poisson-Boltzmann method[10;11] in which the calculation is performed by mapping the electrostatic potential, charge density and dielectric constant to grid points surrounding the

solute. Points that lie within the solute are given a low dielectric, while those in the solvent are given a high dielectric. The electrostatic free energy of solvation can then be obtained from the work required to move the solute charges from vacuum into the electrostatic potential calculated from the Poisson-Boltzmann equation.

Given the complexity of the Poisson-Boltzmann equation, a number of simplifications have been developed. The Born model[12] gives the electrostatic free energy change for the solvation of a point charge q in a spherical cavity of radius a:

$$\Delta G_{elec} = -\frac{q^2}{2a}\left(1 - \frac{1}{\varepsilon}\right) \tag{2.17}$$

The generalised Born method[13] extends this equation to cavities of any shape:

$$\Delta G_{elec} = -\sum_i \frac{q_i^2}{2R_i}\left(1 - \frac{1}{\varepsilon}\right) - \frac{1}{2}\sum_{ij,i\neq j} \frac{q_i q_j}{f^{\mathrm{GB}}\left(r_{ij}, R_i, R_j\right)}\left(1 - \frac{1}{\varepsilon}\right) \tag{2.18}$$

where i and j are sums over charged atoms, q_i is the charge on i, r_{ij} is the separation of i and j and R_i is the effective Born radius of atom i. The function $f^{\mathrm{GB}}\left(r_{ij}, R_i, R_j\right)$ is given by:

$$f^{\mathrm{GB}}\left(r_{ij}, R_i, R_j\right) = \left[r_{ij}^2 + R_i R_j \exp\left(-\frac{r_{ij}^2}{4R_i R_j}\right)\right]^{\frac{1}{2}} \tag{2.19}$$

A number of schemes for calculating the effective Born radii exist[13–18]

The simplest models for the non-polar terms in Equation 2.14 use a simple solute surface area or volume dependent term for the sum of the dispersion, repulsion and cavitation terms. Floris and Tomasi showed that the integral over charge density volumes for the solute/solvent van der Waals interactions can be replaced

by an integral over the solute surface area.[19;20] This allows the three terms to be separated, and the dispersion and repulsion interactions to be calculated by surface itergrals and the cavitation term to be calculated with an surface area or volume dependent term.

2.2 Statistical Mechanics

Classically, the microscopic state of a system of particles (*microstate*) can be fully described by the momentum and configurational vectors $\mathbf{p}$ and $\mathbf{r}$, which consist of the momenta and positions, in each Cartesian coordinate, of all particles in the system. Together these vectors define the phase space of the system, with a point in phase space given by:

$$\mathbf{X} = (\mathbf{p}, \mathbf{r}) \tag{2.20}$$

The energy of a microstate of the system is obtained from the classical Hamiltonian discussed in section 2.1.1:

$$E(\mathbf{p}, \mathbf{r}) = H(\mathbf{p}, \mathbf{r}) = T(\mathbf{p}) + V(\mathbf{r}) \tag{2.21}$$

For a constant number of particles N, volume V and temperature T, the probability $P(\mathbf{p}, \mathbf{r})$ of finding the system at a certain point in phase space is related to the energy of that microstate by the Boltzmann function:

$$\begin{aligned} P(\mathbf{p}, \mathbf{r}) &\propto e^{-\beta E(\mathbf{p},\mathbf{r})} \\ P(\mathbf{p}, \mathbf{r}) &= Ce^{-\beta E(\mathbf{p},\mathbf{r})} \end{aligned} \tag{2.22}$$

where β is related to the thermodynamic temperature T by $\beta = 1/kT$, and k is the Boltzmann constant. The normalisation constant C in Equation 2.22 can be found by integrating the probability distribution function over all phase space and setting the probability of finding the system at any point in phase space to unity:

$$\begin{aligned} 1 &= C \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\beta E(\mathbf{p},\mathbf{r})} \mathrm{d}\mathbf{p}\mathrm{d}\mathbf{r} \\ C &= \frac{1}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\beta E(\mathbf{p},\mathbf{r})} \mathrm{d}\mathbf{p}\mathrm{d}\mathbf{r}} \\ P(\mathbf{p},\mathbf{r}) &= \frac{e^{-\beta E(\mathbf{p},\mathbf{r})}}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\beta E(\mathbf{p},\mathbf{r})} \mathrm{d}\mathbf{p}\mathrm{d}\mathbf{r}} \end{aligned} \tag{2.23}$$

This probability distribution function defines the canonical ensemble of states available to the system at constant NVT. The denominator on the left hand side of Equation 2.23 is known as the partition function Q of the system:

$$P(\mathbf{p},\mathbf{r}) = \frac{1}{Q} e^{-\beta E(\mathbf{p},\mathbf{r})} \tag{2.24}$$

For any macroscopically observable quantity that can be defined as a function of $\mathbf{p}$ and $\mathbf{r}$, its ensemble average value, or expectation value, can be calculated as the integral, over all phase space, of the product of its function of $\mathbf{p}$ and $\mathbf{r}$ and the probability distribution function of the system:

$$\langle A \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} A(\mathbf{p},\mathbf{r}) P(\mathbf{p},\mathbf{r}) \, \mathrm{d}\mathbf{p}\mathrm{d}\mathbf{r} \tag{2.25}$$

Therefore if one can obtain the partition function of the system, statistical mechanics offers a method of calculating observable quantities for that system.

Unfortunately, for anything other than trivial model systems, the partition function cannot be obtained analytically. In a system of interacting particles, the configurational part of the partition function depends on the positions of all particles, in all positions in phase space, which presents an analytically intractable problem. Molecular simulation offers a way of calculating statistical ensembles numerically, allowing average values to be determined over the simulation trajectory. There are two main methods of generating statistical ensembles through simulation: Monte Carlo methods and molecular dynamics simulation.

The theory underlying these methods, and their technical implementation, will be discussed in the next two sections.

2.3 Monte Carlo Methods

Molecular Monte Carlo techniques rely on the fact that the canonical partition function can be factorised into separate momentum and configurational parts, provided that the potentials which act on the system do not depend on momentum (Note: T in this equation refers to the total kinetic energy, not temperature):

$$\begin{aligned} Q &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\beta E(\mathbf{p},\mathbf{r})}\mathrm{d}\mathbf{p}\mathrm{d}\mathbf{r} \\ &= \int_{-\infty}^{\infty} e^{-\beta T(\mathbf{p})}\mathrm{d}\mathbf{p}\int_{-\infty}^{\infty} e^{-\beta V(\mathbf{r})}\mathrm{d}\mathbf{r} \end{aligned} \tag{2.26}$$

The momentum partition function can then be obtained analytically from ideal

gas laws, or ignored completely if you are interested in properties that do not depend on momentum. The configurational partition function Z thus describes how the system deviates from ideal gas behaviour:

$$Z = \int_{-\infty}^{\infty} e^{-\beta V(\mathbf{r})} \mathrm{d}\mathbf{r} \tag{2.27}$$

A Monte Carlo simulation proceeds from a set of initial coordinates by making a random trial displacement in the coordinates at each step of the simulation these trial moves are either accepted or rejected in such a way that as the system evolves it a correct statistical ensemble. The most common method of accepting or reject trial moves is the Metropolis Criterion:

$$p = \min\left(1.0, \quad \exp\left(-\frac{\Delta E}{kT}\right)\right) \tag{2.28}$$

where p is the probability of accepting the move, ΔE is the change in the energy of the system resulting from the trial move, k is the Boltzmann constant and T the simulation temperature. If the trial move causes a reduction in the energy of the systems, it will always be accepted; if it causes an increase in the energy of the system, it will be accepted with a probability equal to the Boltzmann factor of the resulting energy change.

2.4 Molecular Dynamics Simulation

In a molecular dynamics simulation, the system is propagated from an initial point in phase space by integrating Newton's equations of motion for each particle in the system, subject to the forces obtained from the derivative of the potential at each

point in phase space. A key assumption of this method, the *ergodic hypothesis*, is that the time average of the microstates visited during the simulation is equal to the ensemble average, in the limit of long timescales:

$$\langle A \rangle = \lim_{t_n \to \infty} \frac{1}{t_n} \int_{t=0}^{t_n} A(t) \mathrm{d}t \tag{2.29}$$

2.4.1 Integrating Newton's Laws of Motion

The force $\mathbf{F}$ acting on the particles in the system can be calculated from the derivative of the force field potential $V(\mathbf{r})$ with respect to $\mathbf{r}$:

$$\mathbf{F} = -\frac{\mathrm{d}V(\mathbf{r})}{\mathrm{d}\mathbf{r}} \tag{2.30}$$

The resulting acceleration of particle i is given by Newton's second law of motion:

$$\mathbf{a}_i = \frac{\mathbf{F}_i}{m_i} \tag{2.31}$$

where m_i is the mass of atom i. A true trajectory through phase space is represented by a continuous curve $\mathbf{X}(t)$. In order to simulate this trajectory numerically, it is necessary to integrate the equations of motion through a series of discrete steps of finite length Δt (Figure 2.3). A number of schemes have been developed to perform this integration.

Given a set of starting coordinates $\mathbf{r}$ and velocities $\mathbf{v}$ at time t_0, the change in coordinates with time $\mathbf{r}(t)$ can be approximated by the Taylor expansion of $\mathbf{r}(t)$

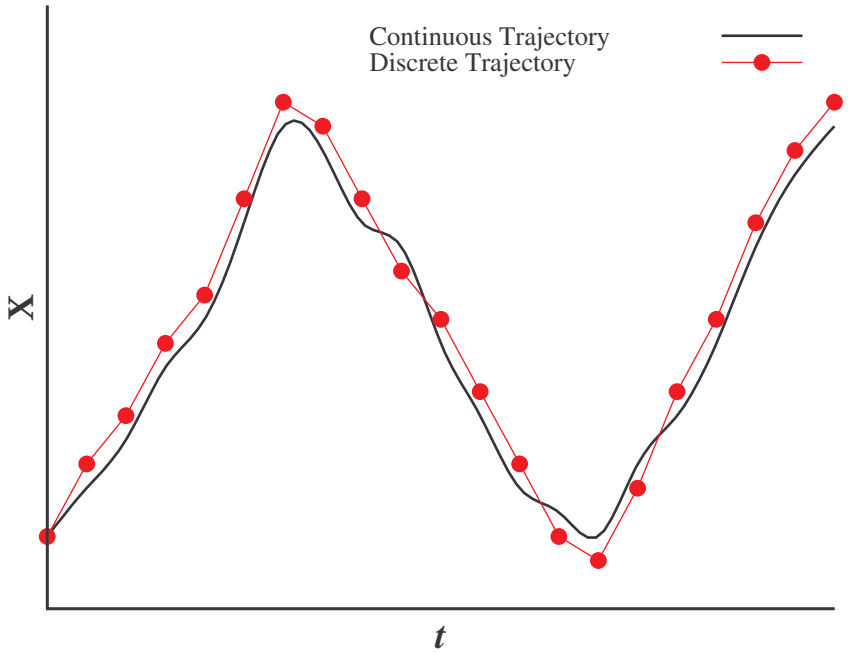


Figure 2.3: Comparison between an actual continuous trajectory through phase space **X** and a numerically integrated discrete trajectory obtained from molecular dynamics simulation.

at t_0:

$$\mathbf{r}(t) = \mathbf{r}(t_0) + (t - t_0)\left.\frac{\mathrm{d}\mathbf{r}(t)}{\mathrm{d}t}\right|_{t=t_0} + \frac{(t-t_0)^2}{2}\left.\frac{\mathrm{d}^2\mathbf{r}(t)}{\mathrm{d}t^2}\right|_{t=t_0} + \frac{(t-t_0)^3}{6}\left.\frac{\mathrm{d}^3\mathbf{r}(t)}{\mathrm{d}t^3}\right|_{t=t_0} + \frac{(t-t_0)^4}{24}\left.\frac{\mathrm{d}^4\mathbf{r}(t)}{\mathrm{d}t^4}\right|_{t=t_0} + \ldots \quad (2.32)$$

Therefore the coordinates at a later time $t_0 + \Delta t$ can be approximated from the coordinates at time t_0 by:

$$\mathbf{r}(t_0 + \Delta t) = \mathbf{r}(t_0) + \Delta t\mathbf{v}(t_0) + \frac{\Delta t^2}{2}\mathbf{a}(t_0) + \frac{\Delta t^3}{6}\dot{\mathbf{a}}(t_0) + \frac{\Delta t^4}{24}\ddot{\mathbf{a}}(t_0) + \ldots \quad (2.33)$$

where we recognise that the first and second derivatives of position at t_0 are equal to the velocity and acceleration at t_0. Similarly, the coordinates at an earlier time

$t_0 - \Delta t$ can be approximated as:

$$\mathbf{r}(t_0 - \Delta t) = \mathbf{r}(t_0) - \Delta t \mathbf{v}(t_0) + \frac{\Delta t^2}{2}\mathbf{a}(t_0) - \frac{\Delta t^3}{6}\dot{\mathbf{a}}(t_0) + \frac{\Delta t^4}{24}\ddot{\mathbf{a}}(t_0) + \ldots \quad (2.34)$$

Through the addition and rearrangement of equations 2.33 and 2.34, and truncating after the Δt^3 term, we arrive at the Verlet algorithm[21;22] for integrating a molecular dynamics trajectory:

$$\mathbf{r}(t_0 + \Delta t) = 2\mathbf{r}(t_0) - \mathbf{r}(t_0 - \Delta t) + \Delta t^2 \mathbf{a}(t_0) \quad (2.35)$$

The Verlet algorithm suffers from two main problems. The velocity does not appear explicitly in the algorithm, making velocity scaling for temperature regulation more difficult (see Section 2.4.3), and the coordinates at two previous time steps are required in order to integrate the current time step. While this may not place a large additional burden on the memory requirements of the algorithm, it does mean that at the first step of the integration, the coordinates $\mathbf{r}(t_0 - \Delta t)$ must be estimated.

Using the Verlet scheme, if velocities at t_0 are required, they can be estimated from:

$$\mathbf{v}(t_0) = \frac{\mathbf{r}(t_0 + \Delta t) - \mathbf{r}(t_0 - \Delta t)}{2\Delta t} \quad (2.36)$$

A modified version of the Verlet algorithm which does include velocity explicitly within the integration scheme is the Leapfrog algorithm. The velocities a half time step after t_0, and a half time step before t_0, can be defined, in an equivalent

manner to Equation 2.36, by:

$$\mathbf{v}\left(t_0 + \frac{\Delta t}{2}\right) = \frac{\mathbf{r}\left(t_0 + \Delta t\right) - \mathbf{r}\left(t_0\right)}{\Delta t} \tag{2.37}$$

$$\mathbf{v}\left(t_0 - \frac{\Delta t}{2}\right) = \frac{\mathbf{r}\left(t_0\right) - \mathbf{r}\left(t_0 - \Delta t\right)}{\Delta t} \tag{2.38}$$

which can be rearranged to give:

$$\mathbf{r}\left(t_0 + \Delta t\right) = \mathbf{r}\left(t_0\right) + \Delta t \mathbf{v}\left(t_0 + \frac{\Delta t}{2}\right) \tag{2.39}$$

$$\mathbf{r}\left(t_0 - \Delta t\right) = \mathbf{r}\left(t_0\right) - \Delta t \mathbf{v}\left(t_0 - \frac{\Delta t}{2}\right) \tag{2.40}$$

where the positions at time $t_0 + \Delta t$ are obtained, via Equation 2.39, using the positions at time t_0 and the velocities at an intermediate time $t_0 + \frac{\Delta t}{2}$. The updates in velocity are found by substituting equations 2.39 and 2.40 into Equation 2.35, and rearranging to give:

$$\mathbf{v}\left(t_0 + \frac{\Delta t}{2}\right) = \mathbf{v}\left(t_0 - \frac{\Delta t}{2}\right) + \Delta t \mathbf{a}\left(t_0\right) \tag{2.41}$$

One drawback to the Leapfrog method is that position and velocity updates are out of phase by half a time step. This means that any calculations of potential and kinetic energy will be out of phase, and so the true total energy of the system cannot be obtained. One way around this problem is to interpolate the velocities to integer time steps.

An alternative Verlet-based method, in which position and velocity updates are synchronised, is the Velocity Verlet algorithm. Rearranging Equation 2.36 gives

$$\mathbf{r}\left(t_0 - \Delta t\right) = \mathbf{r}\left(t_0 + \Delta t\right) - 2\Delta t \mathbf{v}\left(t_0\right) \tag{2.42}$$

substituting Equation 2.42 into the Verlet algorithm (Equation 2.35), and rearranging gives:

$$\mathbf{r}(t_0 + \Delta t) = \mathbf{r}(t_0) + \Delta t \mathbf{v}(t_0) + \Delta t^2 \mathbf{a}(t_0) \tag{2.43}$$

which describes the position updates within the Velocity Verlet scheme. Iterating Equation 2.36 to the next time step gives:

$$\mathbf{v}(t_0 + \Delta t) = \frac{\mathbf{r}(t_0 + 2\Delta t) - \mathbf{r}(t_0)}{2\Delta t} \tag{2.44}$$

then substituting Equation 2.43 for $\mathbf{r}(t_0 + 2\Delta t)$ and $\mathbf{r}(t_0 + \Delta t)$ into Equation 2.44 gives:

$$\mathbf{v}(t_0 + \Delta t) = \mathbf{v}(t_0) + \frac{\Delta t}{2}(\mathbf{a}(t_0) + \mathbf{a}(t_0 + \Delta t)) \tag{2.45}$$

which describes the velocity updates within the Velocity Verlet method.

2.4.2 Periodic Boundary Conditions and Long-Range Interactions

When performing a simulation with an explicit representation of solvent, the computational cost will be dependent on the size of the system under study. It is obviously not possible to simulate a system the size of a macroscopic container, so a microscopic representation is required instead. The effects of the walls of a macroscopic container on the chemistry within will be small, but edge effects can play a significant role when studying a small system. This means the microscopic system being modelled may not represent the macroscopic reality.

One way to avoid these simulation artefacts is to represent the system as a periodic crystal cell. The edges of the simulation box are assumed to be edges of a

unit cell in such a crystal. When one particle leaves the box in one direction, an identical particle will enter from the other direction. Similarly, pair interactions are calculated between particles in adjacent boxes. While this can help avoid edge effect artefacts, it can introduce additional artefacts due to the unnatural periodicity that is added to the system. To avoid these artefacts, the non-bonded interactions between particles are often truncated at a certain distance. By adopting an approach where the distance at which truncation occurs is no more than half the shortest dimension of the periodic simulation box, one can ensure that each particle in the system only feels the direct effects of one copy of each of the other particles. This is known as the *minimum image convention*.

Dispersion interactions vary as r^{-6}, where r is the separation between two interacting particles. The interactions therefore fall rapidly with increasing separations, and so provided the truncation distance is large enough, ignoring interactions past a certain distance has a negligible effect on the dynamics of the system. The interactions between point charges vary as r^{-1}, and so fall far less rapidly with increasing separations. Truncating these interactions can therefore lead to discontinuities in the dynamics of the system, which can effect simulation stability. The Ewald summation,[23] and its derived method particle-mesh Ewald,[24] allows the exact calculation of long range electrostatic interactions in an infinite system. In this method, the charges in the system are neutralised by screening charges described by Gaussian functions. Interactions are then calculated by summing over each charge in the simulation box. Because of the screening charges, these charge interactions are now short-ranged. Added to this is a second summation, performed in reciprocal space from the Fourier transform of the neutralising charge distribution.

2.4.3 Temperature and Pressure Regulation

Following the integration schemes outlined in section 2.4.1 a molecular dynamics simulation naturally conserves energy. Given a fixed simulation box size and number of particles, a molecular dynamics simulation will therefore sample the constant system size, constant volume constant energy (NVE) microcanonical ensemble. It is often desirable to simulate alternative ensembles, and a number of schemes have been developed to allow this.

In the Andersen thermostat,[25] the velocities of a set of particles in the system are periodically reassigned according to the Maxwell-Boltzmann distribution for the desired temperature. In the Berendsen weak coupling method,[26] the velocities of the particles in the system are scaled at each time step, so that over the course of the simulation, the temperature oscillates around the desired value. Both of these methods can be extended to constant pressure regulation, by varying the volume of the simulation box to give the desired simulation pressure.

Langevin dynamics[27] can also be used to control the simulation temperature. Random frictive forces are added to the dynamics of the system which model collisions with particles of an imaginary fluid at a constant temperature. As energy is exchanged with these particles, the temperature of the simulated particles will equilibrate with that of the imaginary particles.

2.5 Advanced Molecular Dynamics Methods

Although MD is a powerful technique for studying the molecular details of biomolecular structure and motion, it is widely recognised that the technique struggles

to adequately sample the rough potential energy surfaces of flexible biomolecules and their complexes. These potential energy surfaces are characterised by a broad range of barriers, at many scales both lower and higher than that of thermal energy kT. Although reduced models, such as lattice or bead approaches, can be applied to relieve this problem, important details can be lost in coarse-graining. Alternatively some approaches alter the energy surface topology, such as the accelerated molecular dynamics[28] and metadynamics[29] methods; alternatively, some approaches modify the sampling regime, for example, the locally enhanced sampling[30] and replica exchange MD approaches[31].

We highlight in a little more detail here two of the most widely used approaches: the metadynamics[29] method is built on the principles of the local elevation technique,[32] applying a history-dependent bias potential during MD simulation to bias the system against returning to previously sampled regions. At the simulation's conclusion, the free energy surface can be obtained from the negative of this potential. An interesting recent application of metadynamics has been to molecular docking. Gervasio et al[33] applied the method to dock a set of four protein–ligand complexes: trypsin/benzamidine, trypsin/chlorobenzamidine, immunoglobulin McPC603/phosphorylcholine and cyclin-dependent kinase 2/staurosporine. Metadynamics simulations based on the AMBER force field[34] and an explicit solvent model were able to recover the correct crystallographic pose from an initial unbound protein/ligand state. In addition, they gave a prediction of the free energy of binding which compared well with estimates from umbrella sampling. Indeed, an additional advantage of metadynamics is not only the information about bound and unbound conformations obtained, but also information on the free energy of intermediates and barriers on the binding pathways.

A second approach which can be used to enhance sampling is the replica exchange

molecular dynamics (REMD) method,[31] also known as also known as parallel tempering. Here, simultaneous independent simulations of a system are performed with each simulation replica at a different temperature. Periodically, exchanges of conformations from higher to lower temperature replicas are attempted, with the probability of an attempted transition being sucessful governed by the Metropolis criterion.[35] This states that when the transition results in a lowering of the potential energy of the system, the attempted exchange is always successful; if the transition would result in an increase in the potential energy of the system, it is accepted with a probability equal to the Boltzmann factor of the resulting energy change. Once transitions are made, the velocities of the replicas are scaled according to the relative temperatures.

In order for efficient conformational sampling to occur, the overlap between the distributions of the potential energy in replicas adjacent to each other in temperature space must be sufficiently high so as to produce a reasonable rate of conformational transitions. As the size of the system under study increases, adjacent replicas must get closer in temperature in order to achieve this overlap. Therefore a larger number of replicas is required to span a given temperature range. Also, when applied to biomolecules such as proteins, the presence of high temperature replicas can cause denaturing of the molecule to occur, moving the simulation away from the low energy conformations of interest. For this reason, simulations of proteins often apply restraints[36] or use a rather limited temperature range.[37]

A number of Hamiltonian replica exchange (HREMD) methods[38–40] have also been reported. In these variants of REMD, independent and non-interacting replicas are simulated at a range of Hamiltonians rather than a range of temperatures. By varying only certain degrees of freedom in the system, a lower number

of replicas of the system are required along with a less demanding computational overhead.

An alternative advanced sampling approach was reported by Huber and van Gunsteren.[41] Their method, named SWARM-MD, uses an interacting swarm of simulation replicas. The implementation and further development of this method forms the basis of Chapter 4 of this work.

2.6 Calculating Binding Free Energies

In order to assess the potential potency of novel inhibitors, accurate methods to predict the binding free energies of ligands with their biomolecular targets are an important tool in rational drug design. While explicit simulation methods such as thermodynamic integration and free energy perturbation[42] offer the potential to accurately obtain these values, intensive calculations are required to get the results to converge to an accurate degree.

The MM–PBSA method[43] offers a way of calculating binding energies by post-processing simulation snapshots. The binding free energy is approximated by:

$$\Delta G_{bind} = G_{complex} - (G_{receptor} + G_{ligand}) \tag{2.46}$$

where:

$$G = \langle E_{MM} \rangle + \langle G_{PBSA} \rangle - TS \tag{2.47}$$

The free energy of each species is estimated from the average force field energy $\langle E_{MM} \rangle$, the average solvation energy $\langle G_{PBSA} \rangle$ and an estimation of the configu-

rational entropy S, which is a measure of the positional freedom of the particles of the system. The first two terms in Equation 2.47 are obtain by averaging over simulation snapshots. Both the complex and free ligand and receptor contributions to Equation 2.46 are often calculated from snapshots arising from a single sinulation of the complex. Excluding electronic effects, the configurational entropy can be separated into translational, rotational and vibrational components (S_{trans}, S_{rot} and S_{vib} respectively), with the total configurational entropy equal to the sum of these terms:

$$S = S_{trans} + S_{rot} + S_{vib} \tag{2.48}$$

The translational entropy can be calculated from ideal gas laws by:[44]

$$S_{trans} = R\left[\ln\left[\left(\frac{2\pi MkT}{h^2}\right)^{\frac{3}{2}}\frac{V^0}{N_A}\right] + \frac{5}{2}\right] \tag{2.49}$$

where R is the gas constant, M is the mass of the species, k is the Boltzmann constant, T is the temperature, h is the Planck constant, V^0 is the standard molar volume, and N_A is Avogadro's number.

The rotational entropy of a non-linear molecule is given by:[44]

$$S_{rot} = R\left[\ln\left[\frac{\sqrt{\pi I_A I_B I_C}}{\sigma}\left(\frac{8\pi^2 kT}{h^2}\right)^{\frac{3}{2}}\right] + \frac{3}{2}\right] \tag{2.50}$$

where σ is a number that relates to rotational symmetry of the species, and I_A, I_B and I_C are the principle moments of inertia of the species. As the moments of inertia of a molecule will change as the conformation changes, the rotational entropy term can be calculated by averaging over a number of snapshots generated from a simulation.

Calculating these first two terms of the configurational entropy is somewhat trivial. The difficulty in these calculations come in the calculation of the vibrational entropy contribution, which is given by:[44]

$$S_{vib} = R \sum_{i=1}^{3N-6} \left[\frac{h\omega_i}{kT\left(e^{h\omega_i/kT} - 1\right)} - \ln\left(1 - \left(e^{-h\omega_i/kT}\right)\right) \right] \tag{2.51}$$

where the sum is over the $3N - 6$ vibrational modes of the molecule (where N is the number of atoms in the molecule). To evaluate this sum, it is necessary to accurately obtain the vibrational frequencies ω_i of each mode.

One way to obtain these frequencies is through normal mode analysis.[45;46] In this method, the species of interest is geometry optimised to a potential energy minimum, and the mass-weighted second derivative Hessian matrix is obtained at this geometry. Diagonalisation of this matrix gives the vibrational normal modes of the species as eigenvectors, and the frequencies of these modes as eigenvalues. The application of this method relies on two main assumptions: firstly, that the true normal modes are harmonic in nature; and secondly, that they are well represented by oscillations around a single potential energy minimum. It is known that large biomolecules cannot be accurately described by a single minimum structure,[47] and that many of their vibrational modes, particulary the low frequency vibrations that produce the largest variations in atomic positions, display a large degree of anharmonicity.[48] Despite these problems, normal mode analysis has been shown to give reasonable entropy estimates, at least when calculating entropy differences.[49]

Quasiharmonic analysis[46;50;51] provides an alternative method for generating vibrational frequencies. Here the frequencies are obtained directly from molecular dynamics simulations. After simulation, the mass-weighted covariance matrix is

obtained from the atomic fluctuations that occur during the simulation. The quasiharmonic frequencies and vibrational modes are then obtained via diagonalisation of this matrix. Unlike normal mode analysis, there is no assumption of harmonicity, therefore quasiharmonic analysis is able to account for the anharmonic nature of low frequency vibrational modes. However, his method does have a number of drawbacks, compared to normal mode analysis. When calculating entropy changes for MM-PBSA analysis of binding free energies, it is necessary to simulate the complex and both unbound species. This can add significantly to the computational cost when the other contributions to the binding free energy are calculated from a single simulation of the complex only. Also, long time scale simulations are required to accurately converge the low frequency vibrational modes.

Chapter 3

Structure and Aggregation Inhibition of β-Amyloid Fibrils

3.1 Aβ Aggregation and its Role in Alzheimer's Disease

Alzheimer's disease (AD) is a progressive neurodegenerative disorder, which involves the gradual onset of symptoms such as confusion, forgetfulness, mood swings and speech impairment. The median life expectancy following diagnosis has been estimated at between 5 – 10 years,[52–55] with pneumonia being the most frequent cause of death.[52;56] It was first described in 1905 by the German neurologist Alois Alzheimer, who identified the presence of unusual tangles and fibrous plaques in the post-mortem examination of the brain of a 56-year old woman who had suffered from progressive dementia over a 5 year period.[57] Over 100 years later, Alzheimer's Disease (AD) affects nearly half a million people in the UK (alzheimers.org.uk) and is expected to double worldwide by 2040.[58] Age and a family history of the disease are major risk factors;[52] the majority of patients are over the age of 65, making up 60% of deaths from dementia.[52]

It transpires that the molecular origins of certain neurodegenerative disorders such as AD, Huntington's,[59] and Parkinson's[60] diseases are associated with the aggregation of proteins.[61] In particular, the accumulation of "plaques" in brain parenchyma and in the walls of cerebral and meningeal blood vessels is typical of AD.[62] These amyloid fibrils are formed through the sequential cleavage of a transmembrane protein, amyloid β precursor protein (APP).[63] APP is cleaved by β-secretase amyloid cleaving enzyme (BACE1) at the N terminal to form a membrane associated APP fragment of 99 amino acids long. Within this 99-residue fragment, a subequent site is cleaved by the enzyme γ-secretase to release a small 39 – 43 amino acid fragment, amyloid β peptide (Aβ).[64] This then accumulates into extracellular fibrillar aggregates and insoluble plaques, consisting of both fib-

rillar and oligomeric Aβ.[65;66] While the 40 residue species Aβ_{40} is the dominant form of Aβ formed within the brain, the 42 residue Aβ_{42} is found to be the major component of aggregated Aβ.[67] Both Aβ_{40} and Aβ_{42} begin with an aspartate residue and terminate with Val40 or Ala42 respectively at the C-terminal end. Within the plaques formed, alterations at the N terminal end can also exist, with the peptides often starting with amino acids Phe3, Ser8 or Glu11. Nevertheless, Aβ_{40} and Aβ_{42} are the major species found. The folding of Aβ in solution is guided by the C-terminal hydrophobic region beginning at residue 29. This directs the carboxy end of the peptide into a β-sheet arrangement. It is thought that the two additional nonpolar residues alanine and isoleucine in Aβ_{42}, increase its aggregation propensity, compared to Aβ_{40}.[68] A second hydrophobic region, Leu17 – Ala21, is known as the central hydrophobic cluster, further stabilises aggregated Aβ.

In healthy individuals, the rate of clearance of Aβ from the brain balances its rate of production. In diseased brains, this balance is disrupted, either by increased production of Aβ or an increase in the proportion of Aβ_{42} compared to Aβ_{40}, leading to increased aggregation producing both soluble Aβ oligomers and protofibrils, and finally the fibrillar deposits that form senile plaques. In addition to this, Aβ aggregation is thought to disrupt the action of kinase and phosphatase enzymes, resulting in an increase in phosphorylated tau. The formation of extremely stable amyloid fibrils provides sinks of Aβ within the brain, preventing its clearance. This description of the progression of AD, known as the Amyloid Cascade Hypothesis,[69] remains the primary theory to explain the pathology of the disease.

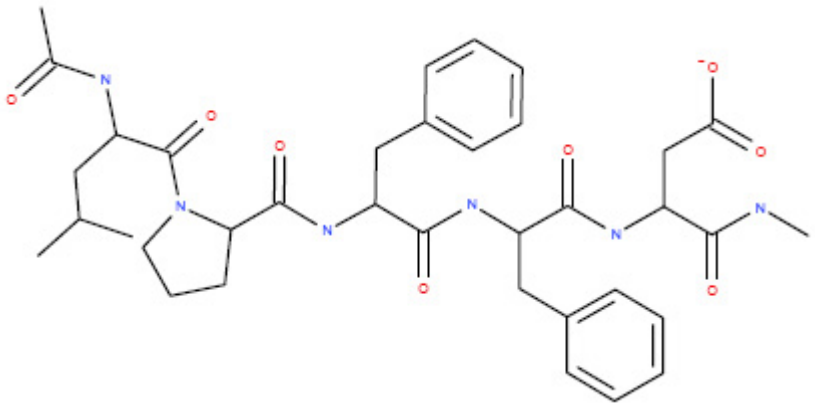
3.2 Peptide Inhibitors of Aβ Aggregation

NOTE: *The work outlined in this section has been published previously in Peptides.*[1]

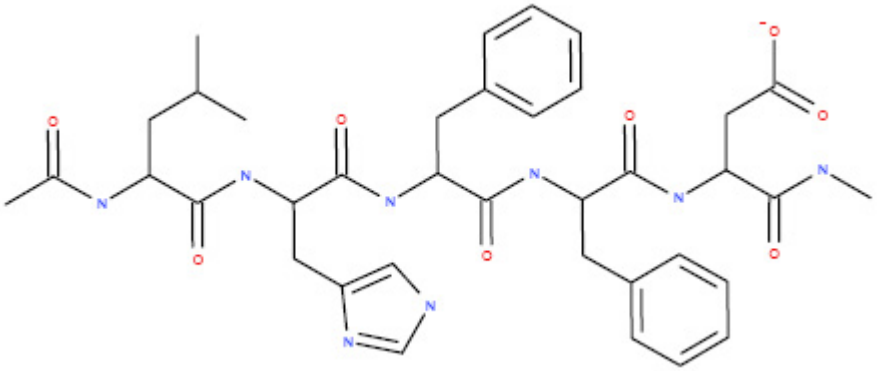
The presence of Aβ aggregates within the brains of Alzheimer's patients, suggests than the identification of small drug-like molecules that can prevent or reverse this aggregation may lead to novel pharmaceutical agents that can halt the progression of the disease. A number of lead compounds have been identified that have been shown to reduce Aβ aggregation, including a series of β-sheet breaker peptides developed by Soto and coworkers.[70] These peptides were designed to mimic the L_{17}VFF hydrophobic region of Aβ, while incorporating a proline residue. It was thought that the presence of the proline residue, with its constrained backbone conformation, would be able to disrupt the β-sheet network in this region of Aβ fibrils. Indeed, it was found that the five residue peptide LPFFD was able to inhibit Aβ aggregation and cause the disassembly of preformed fibrils.[70]

Work performed in the Schein group at the University of Texas Medical Branch sought to identify the structural basis for the activity of these peptides. The inhibitory peptides were docked to an NMR fibril structure. The results were used to design molecular pharmacophores and identify compounds that were shown experimentally to be good inhibitors of aggregation and prevent the neurotoxic effects of Aβ on cultured neuroblastoma cells.[71]

In this section, the stability of the positions predicted by the docking studies[71] was probed using molecular dynamics simulations. Simulations (20 ns in length) were performed of an $A\beta_{9-40}$ 12-mer fibril derived from solid state NMR, complexed with a β-sheet breaker peptide, LPFFD (Figure 3.1a) and with an in-



(a) LPFFD (Ac–Leu–Pro–Phe–Phe–Asp–NMe)



(b) LHFFD (Ac–Leu–His–Phe–Phe–Asp–NMe)

Figure 3.1: Chemical structures of LPFFD and LHFFD.

active control peptide, LHFFD (Figure 3.1b), which lacks the conformationally restricted proline residue, at three discrete lowest energy docking positions. The results of these simulations were used to discern the differences in the ligand–fibril interactions formed by the active and inactive peptides.

3.2.1 Methods

3.2.1.1 Structural Model of Aβ Fibril

The Aβ fibril structural model used in this work was obtained from a series of solid state NMR experiments, performed by Tycko and coworkers, on $A\beta_{40}$ fibrils produced *in vitro*[72;73] (Figure 3.2). The model they obtained omits residues 1 – 8 as these could not be resolved by NMR due to their disordered nature. In this model, the individual $A\beta_{9-40}$ chains form a horseshoe shaped structure consisting of a pair of antiparallel β-sheets separated by a loop region. The 12 $A\beta_{9-40}$ chains present in the model are arranged into two protofibrils (**I** and **II**) each constisting of six chains (**A** – **F**). The protofibrils are stabilised by parallel β-sheet hydrogen bond interactions between adjacent chains, and a row of salt bridges between the Asp23 and Lys28 residues of adjacent chains (which also stabilise the loop regions).

In the discussion that follows, individual residues within the fibril are referred to using the nomenclature $ResX_{A(N)}$, where Res and X are the amino acid residue name and index number within a single chain of Aβ, while A identifies the chains **A** to **F** within each protofibril N.

Figure 3.2 also shows a number of regions within the fibril structure that will be mentioned in the discussion. In addition to the loop regions discussed above, there are two solvent exposed β-sheet faces and two propagation faces to the fibril. The β-sheet faces are formed by the N-terminal residues of the chains of each protofibril; while the propagation faces are perpendicular to the axis along which further Aβ chains attach as aggregation increases.

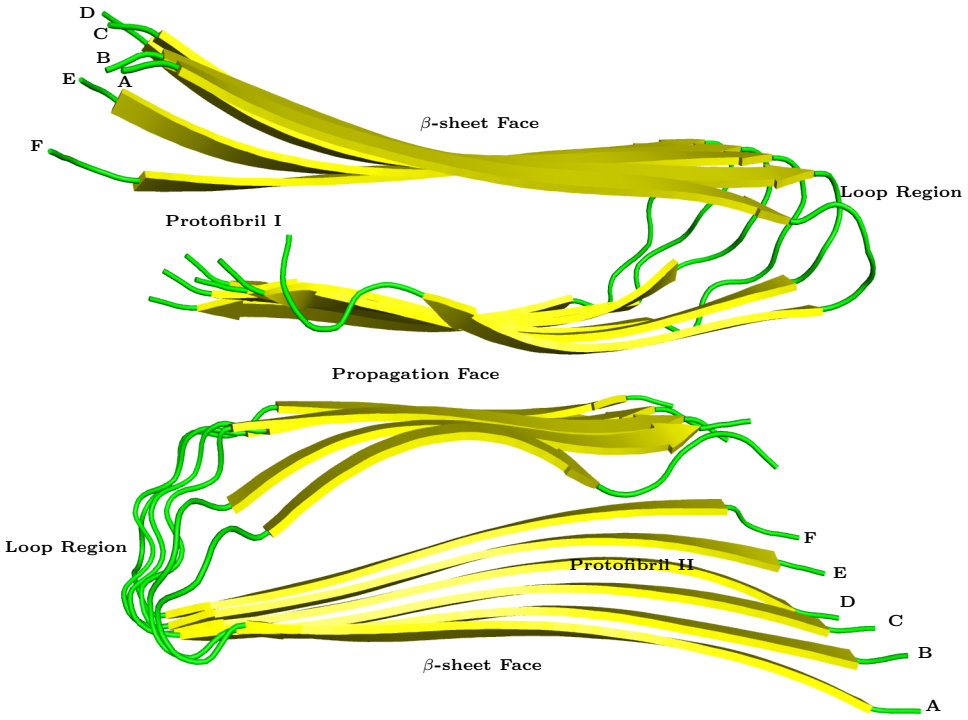


Figure 3.2: Structure of Aβ fibril 12-mer, obtained from Tycko and coworkers.[72;73]

3.2.1.2 Simulation of Fibril–Peptide Complexes

The molecular dynamics trajectories of LPFFD and LHFFD bound to a 12-mer fibrillar aggregate of Aβ were performed by our collaborators at the University of Texas Medical Branch at Galveston, Texas. Throughout the work discussed in this chapter, both LPFFD and LHFFD were modelled with acetyl and methyl amine groups protecting the N- and C-terminal groups, respectively. For both peptides, the C-terminal aspartic acid residue deprotonated, while the histidine residue of LHFFD was singly protonated on the epsilon nitrogen atom. The initial binding positions of the peptides were generated by Chen et al,[71] via molecular docking and scoring with AutoDock,[74;75] using an Aβ fibril structural model provided by Tycko and co-workers[72;73] (Figure 3.2) The top scoring binding site for each peptide was identified, then the second and third highest scoring sites identified, with the constraint that all sites must be no less than 15 Å from each other. Throughout this work, these binding sites are described by their ranking as predicted by AutoDock. As the second and fourth most favourable binding positions of LPFFD, as predicted by AutoDock, where less than 15 Å from the most favourable site (site **1**), the third and fifth highest scoring binding positions were used for further analyses (sites **3** and **5**). Similarly, the position predicted by AutoDock to be the third most favourable site for LHFFD to bind lay less than 15 Å from the most favourable binding position (site **1**), therefore the fourth most favourable position (site **4**) was used in its place during the further analyses.

The simulations and geometry optimisations were performed using *sander* from the Amber 9 molecular modelling software suite,[76] and the Amber *ff03* force field of Duan et al[77] was used. Two simulations were performed: in each case the fibril was complexed with peptide ligands bound in the three predicted binding

sites. The complexes were solvated in a TIP3P[78] water box, using *tleap*,[76] and 12 sodium ions were added to neutralise the periodic simulation cells. For each complex, the water molecules were minimised, with the complex atoms fixed with harmonic restraints of strength 5.0 kcal/mol.Å^2, for 200 steps of steepest descent and 300 steps of conjugate gradient minimisation. The restraints were removed an the simulation box was relaxed by 500 steps of steepest descent and 1500 steps of conjugate gradient minimisation.

During simulation, non-bonded interactions were truncated at 8.0 Å, and particle mesh Ewald[24] was used to model long range electrostatic interactions. A simulation time step of 2.0 fs was used and SHAKE[79] was applied to all bonds involving hydrogen. The simulation temperature was controlled using the Berendsen thermostat[26] with a coupling constant of 1.0 ps. Initial velocities were assigned from a probability distribution at 100 K, and the system was heated over 200 ps to 310 K, with harmonic restraints of strength 5.0 kcal/mol.Å^2 on the atoms of the fibril and inhibitors. The system density was allowed to equilibrate to atmospheric pressure using a weak coupling barostat.[26] The restraints were removed, and the density was allowed to equilibrate for a further period of 200 ps. A production simulation of 20 ns was then performed under constant volume conditions.

3.2.1.3 Structural Analysis of Bound Peptide Poses

The movement of the peptides, relative to the fibril was analysed using *ptraj* from Amber 10.[80] Simulation snapshots were aligned by fibril backbone atoms. Cluster analysis was performed on the position of each peptide within these aligned snapshots, using *kclust* from the MMTSB Tool Set,[81] which employs a modified version of the *k*-means clustering algorithm.[82] The k-means algorithm clusters

each data point, in this case each coordinate snapshot, into k clusters, such that the distance from each data point to the mean of its cluster, measured here as the RMS deviation in atomic position, is minimised. To do this, k cluster centroids are randomly assigned at the start of the algorithm, and each data point is assigned to the cluster with the closest centroid. The positions of the cluster centroids are then updated to the mean value of the data points within their clusters, and the data points are again assigned to the clusters with the closest centroid. These two steps are repeated until all the assignment of data points into clusters remains constant in subsequent iterations. This algorithm does not ensure that the final centroid positions represent the global minimum of the clustering problem, as it is sensitive to the initial choice of randomly assigned centroids. To combat this problem, *kclust* performs an extra iterative step in which the initial choice of cluster centroids is changed, and the algorithm repeated until agreement is found. Additionally, *kclust* repeats the procedure, with increasing numbers of clusters, until the distance between each data point and its cluster centroid is less than a maximum value. This maximum is known as the clustering radius.

Following clustering, the snapshots closest to the centre of each cluster were identified, and visualised to identify interactions between the fibril and ligands. The average interaction distance across all members of each cluster were calculated using *ptraj*.[80]

3.2.1.4 Calculation of Fibril–Peptide Binding Energies

The binding energy free energy of both peptides, in each of their binding positions, was calculated using the MM–PBSA method.[43] A single trajectory approach was used, with the energies of the fibril–peptide complexes, and the free peptide and

fibril energies, calculated from snapshots of the same simulation. The analysis did not include an estimation of the entropic contributions to the binding free energy, as it was assumed these would be fairly constant given the similar nature of the peptides. The molecular mechanics contributions were calculated with the Amber *ff03* force field,[77] as used in the simulation.

Four different methods (methods **A**, **B**, **C** and **D**—see below) were used for the Poisson-Boltzmann–surface area contributions, and their results were compared. The calculations were performed using the *pbsa* module of Amber 10.[80] In all four methods, dielectric constants of 1.0 for the solute and 80.0 for the solvent were used. In method **A**, the modified Bondi radii[15;83] were used to represent the solute atoms, and a probe radius of 1.4 Å was used to define both the boundary between the solute and solvent dielectic regions, and the solvent accessible surface of the solute. The total non-polar contribution to the solvation free energy of each species was calculated with a solvent accessible surface area (SASA) dependent term with a coefficient of 0.0072 kcal/mol.Å^2

In methods **B**, **C** and **D**, the optimised Amber atomic radii of Tan et al were used,[84] and the non-polar contributions to the solvation free energy were decomposed into separate attractive and repulsive contributions. Method **B** used a 6-12 decomposition scheme[19;85;86] with the dispersion contributions calculated using the surface integral approach of Floris and Tomasi.[19] An effective water density of 1.0 was used for these calculations. The repulsive contributions were calculated with a solvent accessible surface area dependent term with a coefficient of 0.04356 kcal/mol.Å^2 and an offset of -1.008 kcal/mol. The solvent accessible surface generated with a probe radius of 1.6 Å was used for the solute/solvent dielectric boundary, the dispersion integration surface, and for the area dependent repulsion term.

Methods **C** and **D** both involved the σ decomposition scheme described by Tan et al,[86] but method **C** used a volume dependent repulsion term, while method **D** used an area dependent term. Both methods used a solvent accessible surface, with a probe radius of 1.6Å, to define the solute/solvent boundary, and a solvent accessible surface, with a probe radius of 0.557 Å, for the dispersion integration surface. The surface integrations were performed with an effective water density of 1.129. In method **C**, a the volume bounded by a solvent accessible surface with probe radius 1.300 Å was used for the volume dependent repulsion term, with a coefficient of 0.0378 kcal/mol.Å and an offset of -0.5692 kcal/mol. Method **D** used a solvent accessible surface, with a probe radius of 0.28 Å, and a coefficient of 0.0894 kcal/mol.Å and an offset of -0.8824 kcal/mol, for the area dependent repulsion term.

3.2.1.5 Structural Analysis of Free Peptides

The conformations exhibited by free LPFFD and LHFFD in aqueous solution were examined through replica exchange simulations.[31] The Amber *ff03* force field[77] was used, and the generalised Born method of Hawkins, Cramer and Truhlar[14;17;18] was employed to represent the solvent environment. Fully extended ($\phi = \psi = 180°$, except proline: $\phi = -61.5°$,$\psi = -176.5°$) LPFFD and LHFFD structures were generated using *tleap*,[76] and geometry optimised in *sander*[76] with 500 steps of steepest descent and 1500 steps of conjugate gradient minimisation to remove any initial bad contacts in the generated structures.

All simulations were performed with *sander*,[76] with a simulation time step of 1.0 fs. SHAKE[79] constraints were applied to bonds involving hydrogen, and non-bonded interactions were not truncated at any distance. Simulation temperatures

were controlled with Langevin dynamics[27] and a collision frequency of 5.0 ps^{-1}. Initial velocities were assigned from a Maxwell-Boltzmann distribution at 30 K, and the peptides were heated to their target temperatures over 60 ps.

The replica temperatures were chosen by first running a series of 4 ns two replica simulations of LPFFD, with exchange attempts every 2 ps. In these simulations the lower temperature replica was at 310 K and the other at a range of temperatures between 330 and 400 K. The replica at 362 K was found to give a transition ratio of 0.21 and the temperatures of the other replica used in the full simulations the were chosen by a geometric progression,[87] such that the temperature of each replica follows:

$$T_{n+1} = \frac{T_n^2}{T_{n-1}} \tag{3.1}$$

The 100 ns replica exchange simulations were performed with eight replicas at temperature of 265, 310, 362, 423, 494, 576, 673 and 786 K. Temperature exchanges were attempted at 2 ps intervals. The conformations of LPFFD and LHFFD were examined in 45000 equally spaced snapshots from the final 90 ns of simulation. These structures were aligned using *ptraj*,[80] then clustered using *kclust*.[81]

3.2.1.6 Computational Resources

The two 20 ns simulations of LPFFD and LHFFD in complex with Aβ, performed by our collaborators in Texas, were run on 8 processor cores, and each required approximately 1200 hours of simulation walltime. MM-PBSA analysis of the binding free energy of each ligand binding pose took approximately 24 hours on a single processor core. In total, six poses were analysed using four MM-PBSA methods, therefore the total CPU time required was around 24 days. The 100

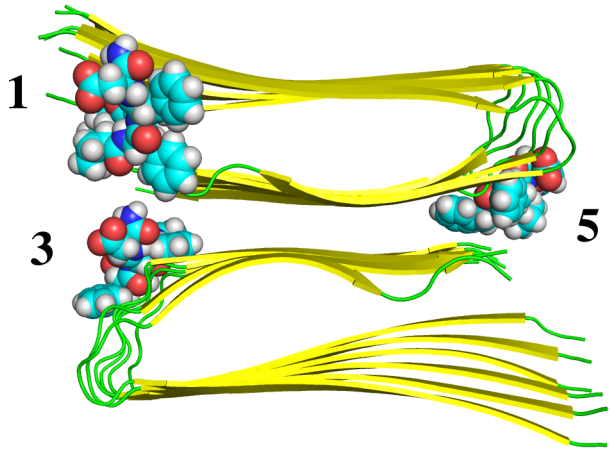
ns replica exchange simulations of free LPFFD and LHFFD in solution took approximately five hours each on 8 processor cores.

3.2.2 Results

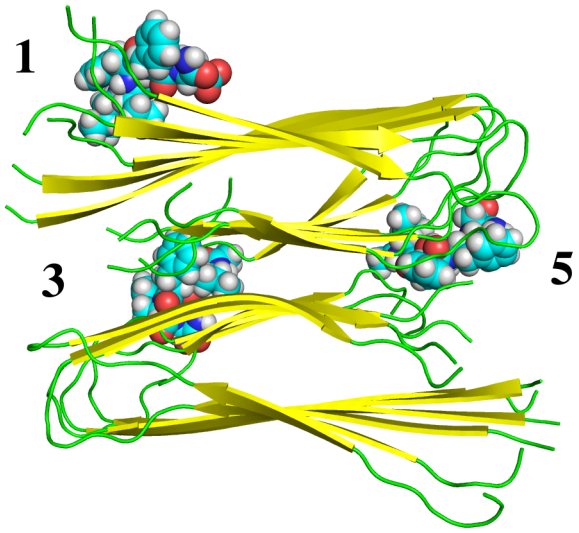
3.2.2.1 Stability of Docked Binding Poses

The stability of the AutoDock predicted binding sites of the active peptide LPFFD, and its inactive analogue LHFFD, was assessed by refinement of the docked poses through 20 ns of molecular dynamics simulation. Interestingly, position **1**, predicted to be the most favourable binding site by AutoDock, was found to be unstable during the simulation. The peptide rapidly detached from its initial position near the N- and C-termini of protofibril **I** (Figure 3.3a), on the propagation face of the fibril, and migrated onto the adjacent β-sheet face (Figure 3.3b). Unsurprisingly, this resulted in a large deviation in the atomic positions of LPFFD, relative to the docked pose, with the RMSD of this peptide rising to approximately 30 Å during the simulation, before settling in a position approximately 20 Å from its initial location (Figure 3.4).

In positions **3** and **5**, LPFFD remained much closer to the AutoDock predicted positions during the simulation. In position **3**, AutoDock predicted a pose near chains **F** and **E** of the C-termini of protofibril **I** and the loop region of protofibril **II** (Figure 3.3a). During the simulation, a small unfolding of the C-terminus of protofibril **I** caused the peptide to move slightly towards the centre of the fibril (Figure 3.3b). The LPFFD peptide in position **5** displayed the smallest displacement from its docked position, following molecular dynamics refinement (Figure 3.4). Throughout the simulation, it remained in the cavity between the



(a) Docked positions



(b) Positions following 20 ns MD

Figure 3.3: Comparison of LPFFD positions following docking and molecular dynamics refinement

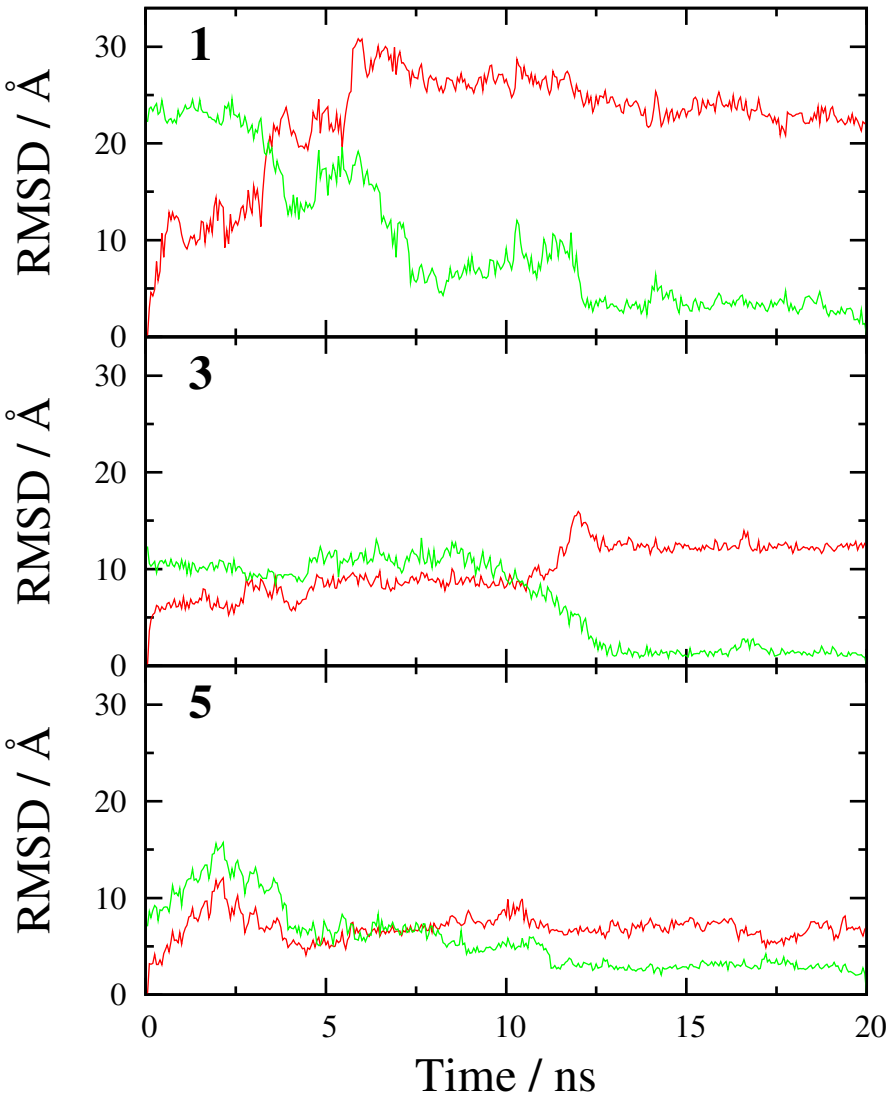


Figure 3.4: RMSD of LPFFD at sites **1**, **3** and **5** relative to the fibril, calculated using the docked pose (red) and final simulation snapshot (green) as reference structures.

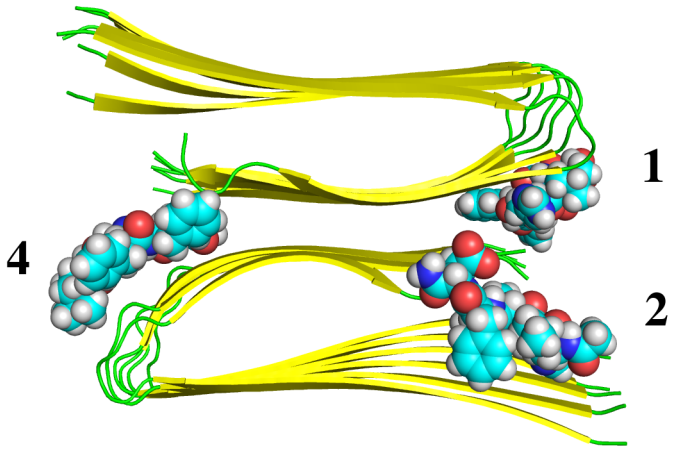
loop region of protofibril **I** and the C-termini of protofibril **II** (Figure 3.3).

Each of three LPFFD peptides achieved stable bound poses by the end of the 20 ns simulation. The RMSD of the peptides, relative to their final molecular dynamics positions, stayed below 5 Å during the final 7.5 ns of simulation (Figure 3.4).

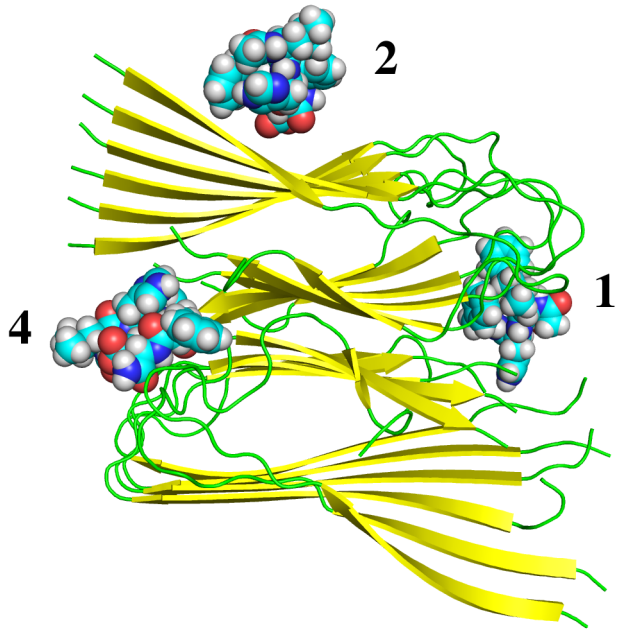
For LHFFD, position **1**, predicted by AutoDock to be the most favourable binding site, lay in the cavity between the loop region of protofibril **I** and the C-termini of protofibril **II** (Figure 3.5a). Throughout the simulation, the peptide remained within 10 Å of this position (Figure 3.6), however LHFFD did not form a stable binding pose. Instead the peptide was observed to traverse along the cavity within the fibril.

The LHFFD peptide bound in position **2**, on the propagation face of the fibril (Figure 3.5a), was found to show no affinity for the fibril during the simulation. Instead it immediately disassociated into the solvent, before reattaching to the periodic image of the fibril, on the β-sheet face of protofibril **I** (Figure 3.5b).

Position **4**, which was predicted by AutoDock to be the least favourable of the three binding sites for LHFFD, lay between the loop region of chain **A** of protofibril **II** and the C-terminus of chain **A** of protofibril **I** (Figure 3.5a). The LHFFD peptide bound in this position showed the least movement during the simulation (Figure 3.5b), with its RMSD staying below 10 Å, relative to its initial position. In the early stages of the simulation, the peptide buried itself deeper into the cavity between the two protofibrils (Figure 3.5b), where it traversed slightly along the groove, between chains **A** and **C**.



(a) Docked positions



(b) Positions following MD

Figure 3.5: Comparison of LHFFD positions following docking and molecular dynamics refinement

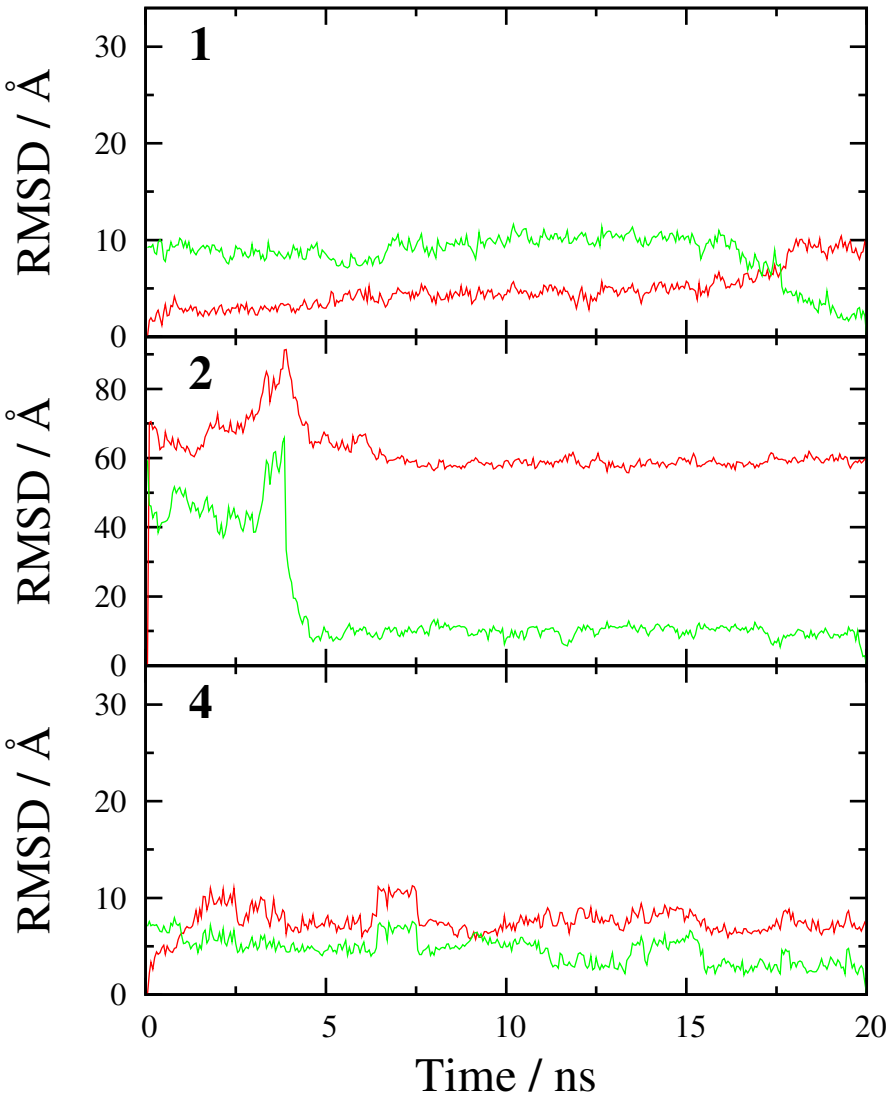


Figure 3.6: RMSD of LHFFD at sites **1**, **2** and **4** relative to the fibril, calculated using the docked pose (red) and final simulation snapshot (green) as reference structures (Note a different y-axis scale is used for position **2**).

3.2.2.2 MM–PBSA Analysis of Fibril–Peptide Binding Energies

The binding affinities of LPFFD and LHFFD to the fibril, at each binding position, were calculated using four different MM–PBSA schemes (Table 3.1). Method **A** predicted that LPPFD bound to the fibril most strongly in position **5**, with a binding free energy of -37.5 ± 6.0 kcal/mol. When using methods **B**, **C** or **D** for MM–PBSA analysis, position **1** was predicted to be the most favourable site for binding, with predicted binding affinities of -27.9 ± 10.3, -11.2 ± 10.9 and -46.7 ± 12.2 kcal/mol respectively. The binding affinity of LPFFD at site **1** was calculated to be -34.0 ± 7.3 kcal/mol by method **A**, slightly lower than that of position **5**.

All four methods agree that position **2** is the most favourable site for LHFFD to bind to the fibril. The calculated binding energies were -27.3 ± 5.6, -24.6 ± 7.7, -40.4 ± 9.3 and -12.8 ± 7.8 kcal/mol when calculated using methods **A**, **B**, **C** and **D** respectively.

Table 3.1: Fibril–peptide binding energies calculated using MM–PBSA with standard deviations. Calculations performed using three different PBSA methods (see section 3.2.1.4).

Ligand	[a] ΔG_{bind}	[b] ΔG_{elec}	[c] ΔG_{vdW}	[d] ΔG_{PB}	[e] ΔG_{rep}	[f] ΔG_{dis}	[g] ΔG_{np}
Method A							
LPFFD (**1**)	−34.0 ± 7.3	−94.8 ± 21.5	−31.9 ± 3.8	98.1 ± 17.6	na	na	−5.5 ± 1.1
LPFFD (**3**)	−27.0 ± 3.9	79.9 ± 24.2	−37.1 ± 3.4	−63.8 ± 23.0	na	na	−6.0 ± 1.7
LPFFD (**5**)	−37.5 ± 6.0	126.4 ± 19.5	−49.2 ± 4.2	−106.0 ± 16.3	na	na	−8.0 ± 1.2
LHFFD (**1**)	−26.2 ± 7.4	105.8 ± 20.7	−41.4 ± 7.5	−84.3 ± 20.9	na	na	−6.3 ± 2.2
LHFFD (**2**)	−27.3 ± 5.6	−69.7 ± 21.0	−22.3 ± 4.1	69.1 ± 17.8	na	na	−4.4 ± 1.1
LHFFD (**4**)	−24.5 ± 5.1	194.6 ± 11.6	−37.5 ± 4.3	−175.8 ± 12.3	na	na	−5.7 ± 1.2
Method B							
LPFFD (**1**)	−27.9 ± 10.3	−94.8 ± 21.5	−31.9 ± 3.8	95.9 ± 16.3	−35.0 ± 2.5	38.0 ± 1.7	2.9 ± 1.2
LPFFD (**3**)	−16.9 ± 5.4	79.9 ± 24.2	−37.1 ± 3.4	−59.7 ± 23.3	−38.7 ± 2.0	38.7 ± 1.5	0.0 ± 1.1
LPFFD (**5**)	−22.1 ± 7.0	126.4 ± 19.5	−49.2 ± 4.2	−98.7 ± 16.2	−52.6 ± 2.0	52.0 ± 1.4	−0.6 ± 1.3
LHFFD (**1**)	−9.5 ± 9.2	105.8 ± 20.7	−41.4 ± 7.5	−76.8 ± 23.3	−39.9 ± 3.8	42.8 ± 3.8	2.9 ± 1.4
LHFFD (**2**)	−24.6 ± 7.7	−69.7 ± 21.0	−22.3 ± 4.1	66.0 ± 15.7	−27.6 ± 3.2	29.1 ± 2.0	1.5 ± 1.3
LHFFD (**4**)	−15.2 ± 8.1	194.6 ± 11.6	−37.5 ± 4.3	−175.0 ± 14.0	−36.7 ± 3.6	39.4 ± 3.8	2.8 ± 1.5

Continued on next page

Table 3.1 – continued from previous page

Ligand	[a] ΔG_{bind}	[b] ΔG_{elec}	[c] ΔG_{vdW}	[d] ΔG_{PB}	[e] ΔG_{rep}	[f] ΔG_{dis}	[g] ΔG_{np}
Method C							
LPFFD (**1**)	−11.2 ± 10.9	−94.8 ± 21.5	−31.9 ± 3.8	95.3 ± 16.2	−25.7 ± 2.3	45.8 ± 2.9	20.2 ± 1.3
LPFFD (**3**)	3.8 ± 5.5	79.9 ± 24.2	−37.1 ± 3.4	−60.1 ± 23.2	−28.2 ± 2.2	49.4 ± 2.8	21.2 ± 1.3
LPFFD (**5**)	7.0 ± 7.1	126.4 ± 19.5	−49.2 ± 4.2	−99.3 ± 16.1	−38.2 ± 2.4	67.2 ± 2.9	29.0 ± 1.4
LHFFD (**1**)	11.1 ± 9.2	105.8 ± 20.7	−41.4 ± 7.5	−77.1 ± 23.1	−29.7 ± 4.0	53.5 ± 6.3	23.8 ± 2.8
LHFFD (**2**)	−12.8 ± 7.8	−69.7 ± 21.0	−22.3 ± 4.1	65.6 ± 15.7	−19.9 ± 2.7	33.5 ± 3.9	13.6 ± 1.8
LHFFD (**4**)	4.1 ± 9.0	194.6 ± 11.6	−37.5 ± 4.3	−175.0 ± 13.9	−26.9 ± 3.1	48.9 ± 4.9	22.0 ± 2.5
Method D							
LPFFD (**1**)	−46.7 ± 12.2	−94.8 ± 21.5	−31.9 ± 3.8	95.3 ± 16.2	−61.2 ± 5.4	45.8 ± 2.9	−15.3 ± 3.0
LPFFD (**3**)	−35.2 ± 6.7	79.9 ± 24.2	−37.1 ± 4.3	−60.1 ± 23.2	−67.3 ± 5.1	49.4 ± 2.8	−17.8 ± 3.0
LPFFD (**5**)	−45.6 ± 8.8	126.4 ± 19.5	−49.2 ± 4.2	−99.3 ± 16.1	−90.7 ± 5.8	67.2 ± 2.9	−23.5 ± 3.5
LHFFD (**1**)	−29.9 ± 11.3	105.8 ± 20.7	−41.4 ± 7.5	−77.1 ± 23.1	−70.6 ± 9.5	53.5 ± 6.3	−17.2 ± 4.1
LHFFD (**2**)	−40.4 ± 9.3	−69.7 ± 21.0	−22.3 ± 4.1	65.6 ± 15.7	−47.5 ± 6.3	33.5 ± 3.9	−14.0 ± 3.1
LHFFD (**4**)	−33.1 ± 9.0	194.6 ± 11.6	−37.5 ± 4.3	−175.0 ± 13.9	−64.1 ± 7.2	48.9 ± 4.9	−15.2 ± 3.4

[a] MM–PBSA binding free energy, excluding entropic contributions.
[b] Electrostatic contribution to the binding free energy.
[c] van der Waals contribution to the binding free energy.
[d] Poisson-Boltzmann electrostatic solvation contribution to the binding free energy.
[e] Non-polar repulsive contribution to the binding free energy.
[f] Non-polar dispersion contribution to the binding free energy.
[g] Total non-polar contribution to the binding free energy.

3.2.2.3 Cluster Analysis of Bound Peptide Conformations

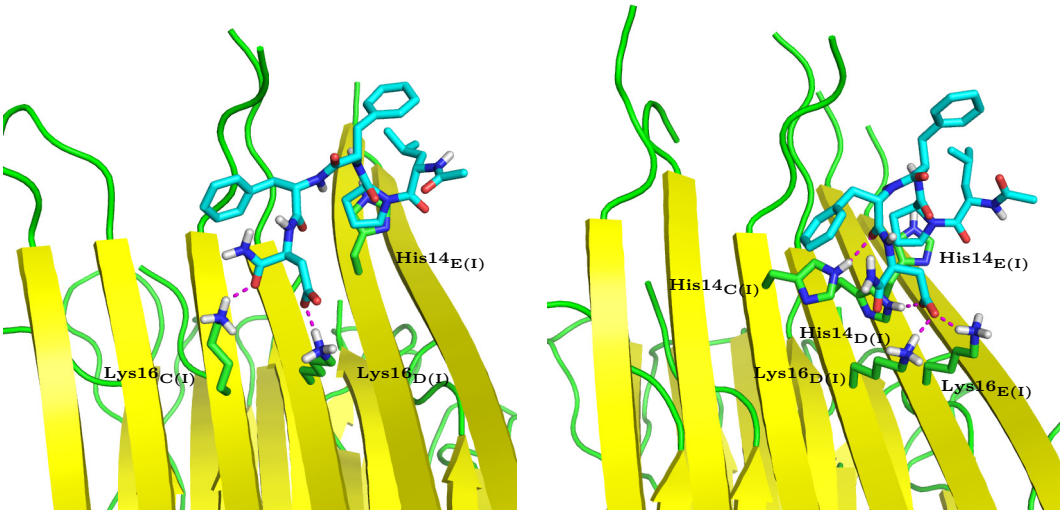
Cluster analysis of the poses adopted by LPFFD when bound to the fibril, during the final 7.5 ns of simulation, was performed using a cluster radius of 1.5 Å. The analysis identified five interaction clusters (**1a** – **1e**) in position **1**, a single cluster (**3a**) in position **3** and four clusters (**5a** – **5d**) in position **5**. The ligand–fibril interactions present in the simulation snapshot closest to each cluster centroid were identified, and the average interaction distances within the cluster were calculated (Table 3.2).

In the representative structures of both of most highly populated clusters at position **1**, which together account for 84% of simulation snapshots, the ligand is stabilised through a salt bridge interaction between the $Lys16_{D(I)}$ side chain of the fibril and the aspartate residue of the ligand (Figure 3.7). Further stabilisation is also provided by a stacking interaction between the $His14_{E(I)}$ residue and the proline present in LPFFD. This stacking interaction appears to be particularly well conserved during the simulation, appearing in all but the least populated cluster, which only accounts for 2% of simulation snapshots (Table 3.2). More transient interactions with His14 and Lys16 residues were also observed (Table 3.2).

The LPFFD ligand in position **3** showed the smallest changes in its coordinates over the final 7.5 ns of the simulation (Figure 3.4). During this period its average atomic RMSD, relative to its final simulation position, was 1.44 Å, with the maximum value in any snapshot 2.78 Å. For this reason, all snapshots were assigned to the same cluster. The only specific fibril-LPFFD interaction which was observed was a salt bridge between the $Lys28_{F(II)}$ side chain of the fibril and the aspartate of LPFFD (Figure 3.8, Table 3.2). Instead, LPFFD appears to be

Table 3.2: Fibril-LPFFD interactions identified from cluster analysis (stacking interactions in bold); average interaction distances (r) (between donor and acceptor heteroatoms for salt bridges and hydrogen bonds; ring centres for stacking interactions) and populations ($\%p$).

Pose/Cluster	Fibril Atoms	Ligand Atoms	r/Å	$\%p$
1a	$Lys16_{D(I)}$ Nζ	Asp5 Oδ1	2.77	53
	$Lys16_{C(I)}$ Nζ	Asp5 O	2.73	
	$His14_{E(I)}$	**Pro2**	**3.97**	
1b	$Lys16_{D(I)}$ Nζ	Asp5 Oδ2	2.93	31
	$Lys16_{E(I)}$ Nζ	Asp5 Oδ2	2.88	
	$His14_{D(I)}$ Nε2	Asp5 Oδ2	2.88	
	$His14_{C(I)}$ Nε2	Phe4 O	3.07	
	$His14_{E(I)}$	**Pro2**	**4.09**	
1c	$Lys16_{C(I)}$ Nζ	Asp5 Oδ1	2.80	11
	$His14_{E(I)}$	**Pro2**	**4.32**	
1d	**$His14_{E(I)}$**	**Pro2**	**4.57**	3
1e	$Lys16_{D(I)}$ Nζ	Asp5 Oδ1	3.09	2
	$Lys16_{D(I)}$ Nζ	Asp5 Oδ2	3.07	
3a	$Lys28_{F(II)}$ Nζ	Asp5 Oδ2	2.91	100
5a	$Ser26_{E(I)}$ Oγ	Asp5 Oδ1	2.69	42
5b	$Ser26_{E(I)}$ Oγ	Asp5 Oδ1	2.64	38
	$Asn27_{D(I)}$ Nδ2	Asp5 Oδ1	2.98	
5c	$Ser26_{E(I)}$ Oγ	Asp5 Oδ1	2.55	17
5d	$Asn27_{D(I)}$ Nδ2	Asp5 Oδ1	2.73	3



(a) Cluster **1a** (53%). **(b)** Cluster **1b** (31%).

Figure 3.7: LPFFD–fibril interactions in the two most populated clusters at position **1**, with cluster occupancy shown in parentheses.

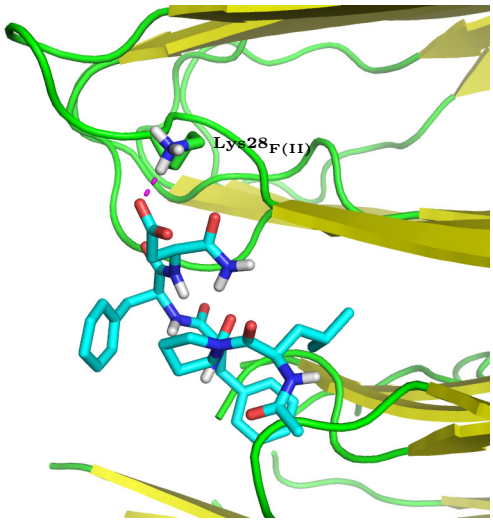
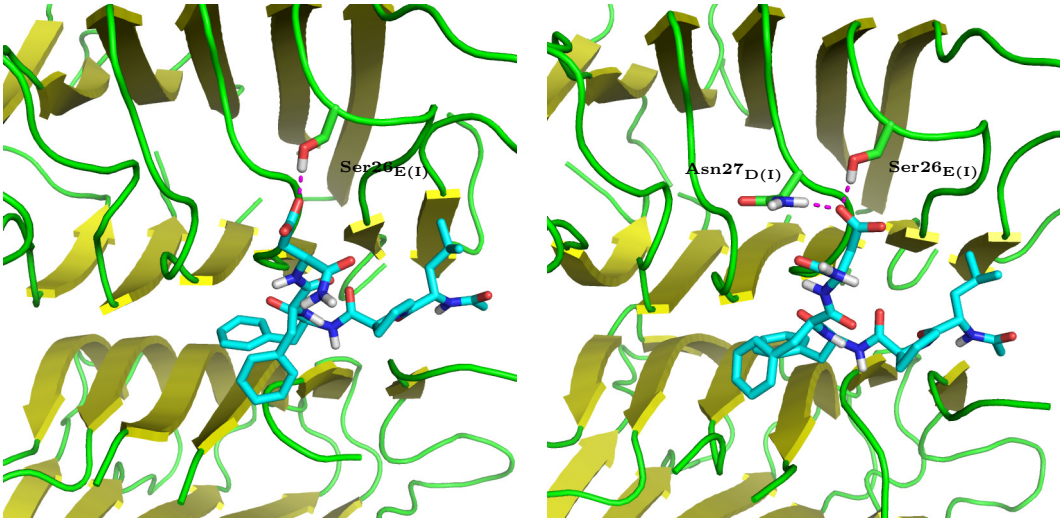


Figure 3.8: LPFFD–fibril interactions in the only cluster **3a** (100% of simulation snapshots) at position **3**.



(a) Cluster **5a** (42%). **(b)** Cluster **5b** (38%).

Figure 3.9: LPFFD–fibril interactions in the two most populated clusters at position **5**, with cluster occupancy shown in parentheses.

stabilised by the burial of its hydrophobic Leu1 and Phe3 residues within the hydrophobic core of the fibril.

Together, the two most populated clusters of LPFFD binding poses at position **5** account for 80% of simulation snapshots (Table 3.2). In the most populated cluster, **5a**, the fibril-LPFFD interaction is stabilised by a hydrogen bond between the $Ser26_{E(I)}$ side chain and aspartate (Figure 3.9a). In cluster **5b** this interaction is also present, with additional stabilisation provided by an additional hydrogen bond between $Asn27_{D(I)}$ and the aspartate of LPFFD (Figure 3.9b). As was observed at position **3**, at position **5** LPFFD appears to further stabilisation by the burial of its hydrophobic residues within the fibril.

As LHFFD displayed a greater variation in atomic RMSD, relative to the final simulation snapshot (Figure 3.6), a slightly larger clustering radius of 2 Å was required to keep the number of clusters to a manageable amount. Using this radius, cluster analysis identified six clusters (**1a** – **1f**) of bound LHFFD poses

at position **1**, five (**2a** – **2e**) at position **2** and six (**4a** – **4f**) at position **4**. The interactions present in these clusters were identified (Table 3.3).

Table 3.3: Fibril-LHFFD interactions identified from cluster analysis; average interaction distances (r) (between donor and acceptor heteroatoms) and populations ($\%p$).

Pose/Cluster	Fibril Atoms	Ligand Atoms	r/Å	$\%p$
1a	$Ser26_{D(I)}$ N	Asp5 Oδ2	2.75	41
1b	No specific interactions			28
1c	$Lys28_{C(I)}$ N	Ace O	3.03	12
1d	$Gly9_{F(II)}$ N	Asp5 Oδ1	2.86	11
	$Lys28_{C(I)}$ O	Leu1 N	3.05	
	$Lys28_{C(I)}$ N	Ace O	2.76	
1e	$Lys28_{C(I)}$ N	Ace O	2.92	7
	$Lys28_{C(I)}$ O	Leu1 N	2.76	
	$Asn27_{D(I)}$ Nδ2	Leu1 N	2.94	
1f	$Val40_{C(II)}$ O	His2 Nε2	2.96	1
2a	$His14_{B(I)}$ Nε2	Asp5 Oδ1	3.17	63
2b	$Lys16_{C(I)}$ Nζ	Asp5 Oδ1	2.72	18
	$Lys16_{B(I)}$ Nζ	Asp5 O	3.04	
	$His14_{B(I)}$ Nε2	Asp5 Oδ2	2.78	
	$His14_{A(I)}$ Nε2	Asp5 O	3.25	
2c	$Lys16_{B(I)}$ Nζ	Asp5 Oδ1	2.89	10
	$His14_{B(I)}$ Nε2	Asp5 Oδ2	2.76	
	$Lys16_{C(I)}$ Nζ	Asp5 Oδ2	3.02	
2d	$Lys16_{C(I)}$ Nζ	Asp5 Oδ1	2.67	7
	$Lys16_{B(I)}$ Nζ	Asp5 Oδ1	2.67	
	$His14_{B(I)}$ Nε2	Asp5 Oδ2	2.91	
2e	$His14_{B(I)}$ Nε2	Asp5 Oδ2	2.62	2
	$Lys16_{D(I)}$ Nζ	Asp5 Oδ2	2.8	
	$Lys16_{B(I)}$ Nζ	Phe4 O	2.86	
	$Lys16_{C(I)}$ Nζ	Asp5 Oδ1	2.99	

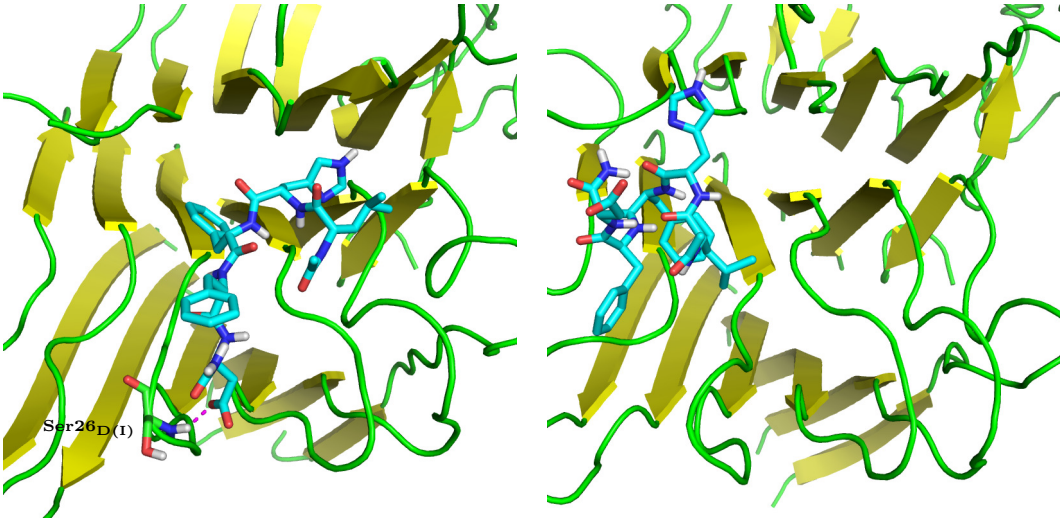
Continued on next page

Table 3.3 – continued from previous page

Pose/Cluster	Fibril Atoms	Ligand Atoms	r/Å	$\%p$
4a	$Asn27_{A(II)}$ Nδ2	Phe3 O	2.81	40
4b	$Asn27_{A(II)}$ Nδ2	Phe3 O	2.78	31
	$Asn27_{B(II)}$ Nδ2	His2 O	2.82	
4c	$Asn27_{A(II)}$ Nδ2	Phe3 O	2.88	17
	$Asn27_{B(II)}$ Nδ2	Asp5 O	3.11	
	$Asn27_{C(II)}$ Nδ2	Ace O	2.85	
4d	$Asn27_{A(II)}$ Nδ2	Phe3 O	3.17	6
4e	$Asn27_{A(II)}$ Nδ2	Phe3 O	3.08	3
4f	$Asn27_{A(II)}$ Nδ2	Phe3 O	2.74	3

In position **1**, the representative structure of the most populated cluster shows LHFFD occupying the cavity between the loop region of protofibril **I** and the C-termini of protofibril **II** in the region of chains **B** to **D** (Figure 3.11a). This position is stabilised by a hydrogen bond between the main chain carbonyl group of $Ser26_{D(I)}$ of the fibril and the Asp5 side chain of LHFFD (Figure 3.11a, Table 3.3). Additional stabilisation is provided by the burial of the Phe3 residue of the peptide within the hydrophobic fibril interior (Figure 3.11a). In the second most populated cluster, LHFFD lies in the cavity between the loop region of chain **D** to **F** of protofibril **I** and the C-termini of chains **E** and **F** on protofibril **II**. No specific fibril–peptide interactions were identified, though burial of the Leu1 and Phe3 residues of LHFFD within the fibril was observed.

The representative structure of the most populated cluster in position **2** shows the fibril–LHFFD complex is stabilised by a hydrogen bond between the Asp5 residue of LHFFD and the His14 residue of chain **B** on protofibril **I**. The peptide lies parallel to the fibril β-sheets, straddling chains **B** and **C** (Figure 3.11a). In

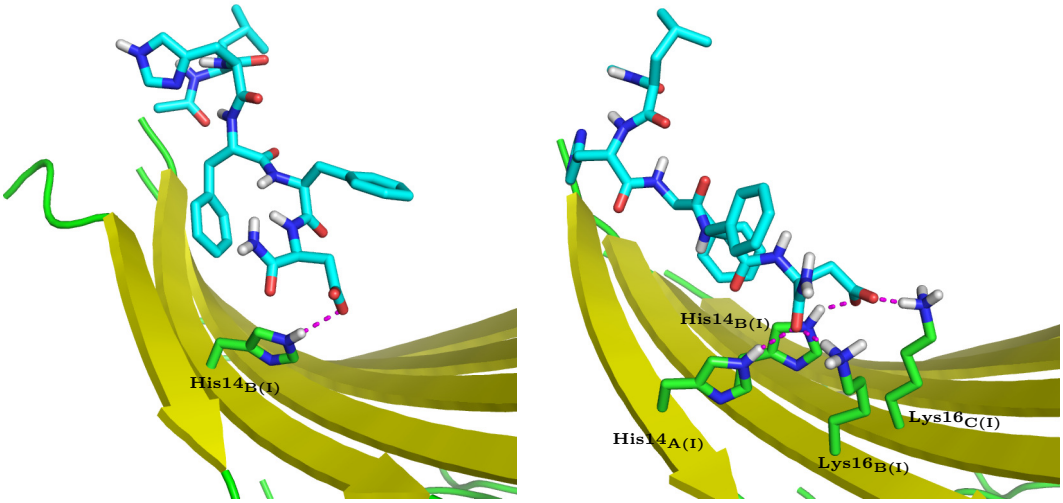


(a) Cluster **1a** (41%). **(b)** Cluster **1b** (28%).

Figure 3.10: LHFFD–fibril interactions in the two most populated clusters at position **1**, with cluster occupancy shown in parentheses.

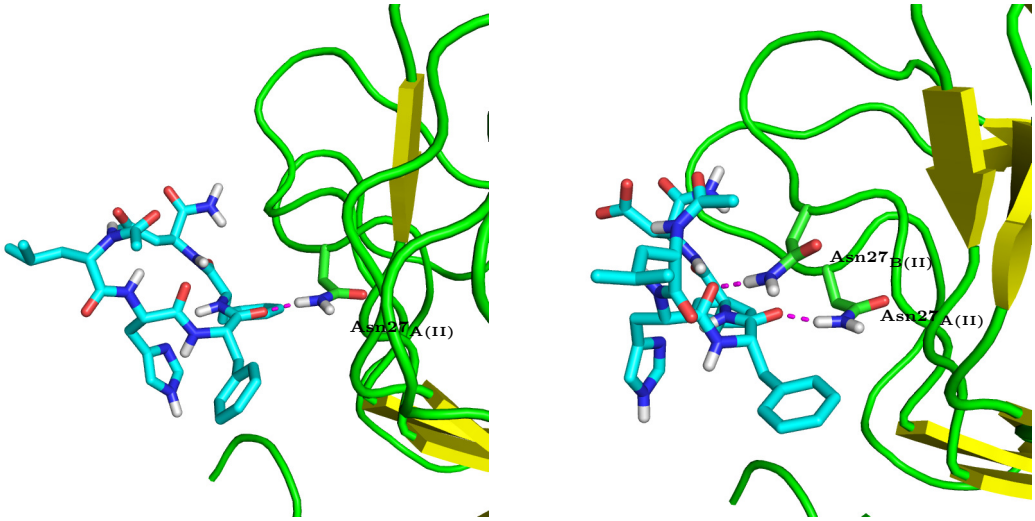
the second most populated cluster, the peptide moved sideways, with its backbone aligned between chains **A** and **B** (Figure 3.11b). The interaction with the $His14_{B(I)}$ residue was still present, and additional interactions with the His14 residue of chain **A** and the Lys16 residues of chain **B** and **C** were observed (Figure 3.11b, Table 3.3).

In the two most populated clusters at position **4**, the fibril–LHFFD interaction is stabilised by the burial of the Phe3 and Phe4 peptide residues within the cavity between the C-termini of chains **A** and **B** of protofibril **I** and the loop region of chains **A** to **C** in protofibril **II** (Figure 3.12). In the most populated cluster **4a**, a hydrogen bond between the main chain carbonyl of Phe3 of LHFFD and the side chain of $Asn27_{B(I)}$ in the fibril was observed (Figure 3.12a, Table 3.3). This hydrogen bond is also present in cluster **4b**, along with an additional hydrogen bond between the Asn27 side chain of chain **B** and the carbonyl oxygen of His2 (Figure 3.12b, Table 3.3).



(a) Cluster **2a** (63%). **(b)** Cluster **2b** (18%).

Figure 3.11: LHFFD–fibril interactions in the two most populated clusters at position **2**, with cluster occupancy shown in parentheses.



(a) Cluster **4a** (40%). **(b)** Cluster **4b** (31%).

Figure 3.12: LHFFD–fibril interactions in the two most populated clusters at position **4**, with cluster occupancy shown in parentheses.

3.2.2.4 Cluster Analysis of Free Peptide Conformations

The conformational flexibility of free LPFFD and LHFFD was investigated through 100 ns replica exchange simulations with eight replicas per peptide, ranging in temperature from 265 to 786 K. The acceptance ratios obtained from the simulations are summarised in Table 3.4.

Table 3.4: Acceptance ratios obtained during 100 ns replica exchange simulations of LPFFD and LHFFD in free solution.

Temperature / K	LPFFD	LHFFD
265 – 310	0.20	0.20
310 – 362	0.21	0.19
362 – 423	0.20	0.19
423 – 494	0.21	0.21
494 – 576	0.23	0.22
576 – 673	0.23	0.22
673 – 786	0.24	0.22

45000 equally spaced snapshots from the final 90 ns of each 310 K simulation were aligned by C_α position, then clustered by heavy atom position using a clustering radius of 3 Å. This clustering identified 14 clusters of LPFFD conformations and 17 clusters of LHFFD conformations. The populations of these clusters are shown in Table 3.5.

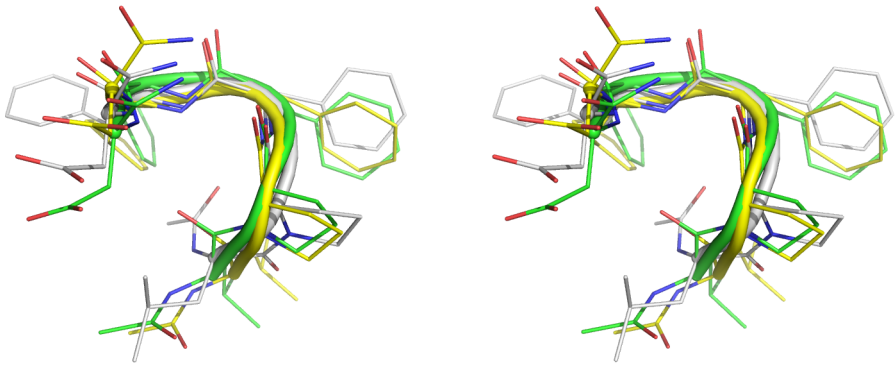
Examining the LPFFD structures closest to each cluster centroid, two backbone conformations were seen to dominate. In clusters **1**, **4** and **6**, which account for 61.4% of LPFFD conformations, the peptide shows a compact turn backbone conformation (Figure 3.13a), with the structural differences between these clusters occuring largely in the amino acid side chain positions. In clusters **2**, **3** and **5**, 32.4% of conformations, LPFFD exhibits a more open backbone conformation (Figure 3.13b). In contrast to LPFFD, a single backbone conformation domi-

Table 3.5: Free peptide conformational clusters and populations (%p).

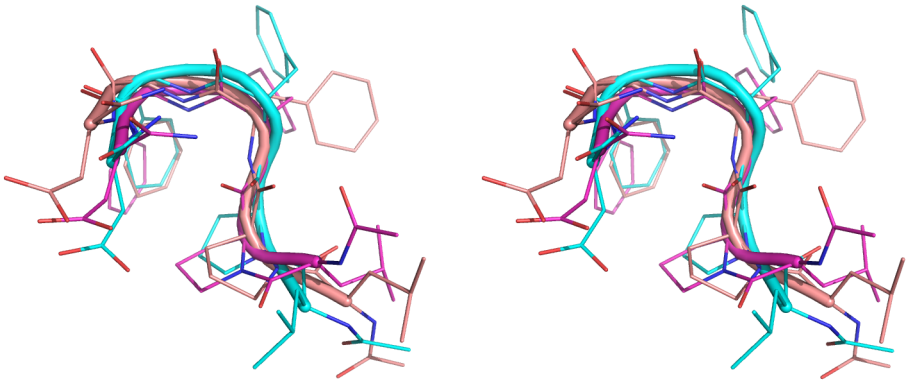
Cluster	LPFFD %p	LHFFD %p
1	45.7	26.2
2	14.4	23.1
3	14.2	17.3
4	12.8	15.5
5	3.8	4.5
6	2.9	2.6
7	1.4	2.3
8	1.4	1.7
9	1.0	1.4
10	0.8	1.3
11	0.7	0.8
12	0.4	0.8
13	0.4	0.6
14	0.2	0.6
15	na	0.6
16	na	0.4
17	na	0.3

nated the ensemble of LHFFD structures obtained from the simulation. Clusters **1** – **5**, accounting for 86.6% of snapshots, all showed a compact turn conformation (Figure 3.14).

The LPFFD and LHFFD conformations obtained from simulations of the peptides free in solution were compared to those observed in the bound state when complexed to the fibril (Tables 3.6 and 3.7). The compact turn conformation shown in clusters **1**, **4** and **6** of LPFFD in free solution does not appear to match any of its bound conformations. Similarly, the bound geometries of LHFFD do not, in general, match those of the compact turn structure found in clusters **1** – **5** in free solution. One exception to this is the bound state of LHFFD in cluster **1b**, which shows a very good agreement with conformations found in the free state. In contrast, clusters **2**, **3** and **5** of free LPFFD are a far better match to the bound conformations in found in the most populated clusters of LPFFD in



(a) LPFFD Clusters **1**, **4** and **6**.



(b) LPFFD Clusters **2**, **3** and **5**.

Figure 3.13: Stereoscopic views of six most populated clusters of LPFFD, obtained from replica exchange simulation of free peptide

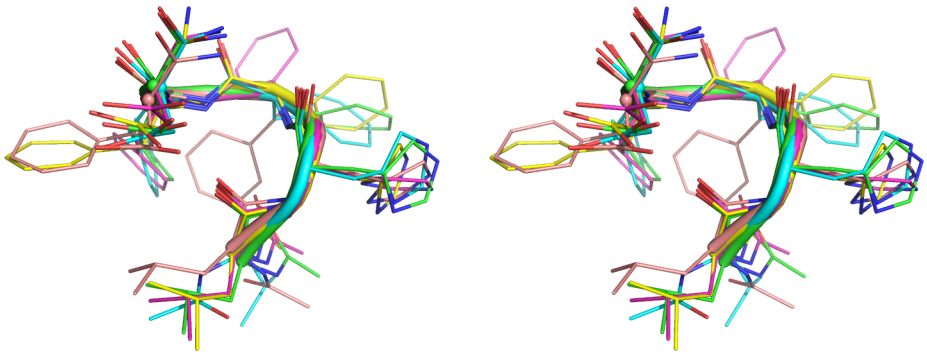


Figure 3.14: Stereoscopic view of five most populated clusters (**1** – **5**) of LHFFD, obtained from replica exchange simulation of free peptide

Table 3.6: RMSD (Å) of C_α atoms between representative cluster structures for LPFFD free in solution and LPFFD bound to the fibril in positions **1**, **3** and **5**.

	1a	**1b**	**3a**	**5a**	**5b**
1	2.35	2.10	1.83	2.15	2.13
2	1.49	0.90	0.69	1.08	0.86
3	1.23	0.53	0.44	1.25	1.01
4	1.93	1.97	1.98	2.02	2.10
5	0.88	0.87	1.20	0.77	0.79
6	2.17	1.99	1.81	2.12	2.10
7	1.91	2.40	2.76	2.13	2.24
8	1.10	0.85	1.06	0.31	0.20
9	1.93	2.23	2.56	1.84	1.90
10	1.70	2.26	2.71	1.97	2.10
11	1.89	2.20	2.62	2.36	2.35
12	1.48	2.09	2.58	1.80	1.94
13	2.07	1.75	1.72	1.85	1.66
14	1.99	1.76	1.60	1.80	1.73

Table 3.7: RMSD (Å) of C_α atoms between representative cluster structures for LHFFD free in solution and LHFFD bound to the fibril in positions **1**, **2** and **4**.

	1a	**1b**	**2a**	**2b**	**4a**	**4b**
1	2.57	0.54	2.59	2.52	1.86	2.08
2	2.68	0.50	2.60	2.59	1.97	2.18
3	2.61	0.36	2.61	2.55	1.99	2.23
4	2.64	0.29	2.65	2.61	1.99	2.26
5	2.59	0.33	2.79	2.64	1.90	2.17
6	1.93	1.06	1.85	1.72	2.25	2.41
7	2.62	1.41	1.85	2.16	2.22	2.13
8	2.21	2.26	2.59	2.62	1.02	0.67
9	0.76	2.37	1.76	1.34	1.97	1.94
10	2.11	1.04	2.99	2.56	1.38	1.77
11	0.94	2.65	1.10	0.80	2.68	2.60
12	0.79	2.36	1.23	0.56	2.50	2.48
13	1.33	2.36	0.86	0.96	2.85	2.83
14	2.22	3.28	1.68	1.48	3.15	2.94
15	2.49	2.41	1.26	1.81	2.56	2.34
16	1.70	2.71	2.00	1.89	2.01	1.74
17	2.14	2.89	1.64	1.40	2.93	2.77

all three positions, the C_α atom RMSD being below 1.5 Å in all cases.

3.2.3 Discussion

3.2.3.1 Assessment of AutoDock Predicted Poses

Performing molecular dynamics refinement following molecular docking allowed us to assess the quality of the bound poses predicted by AutoDock. For both LPFFD and LHFFD, the ligand in the position which AutoDock predicted to be the most favourable for binding that ligand (site **1** in each case) was displaced by a significant distance during simulation (Figures 3.4 and 3.6). In the case of LHFFD, the ligand was seen to move within the cavity between the loop region of protofibril **1** and the C-termini of **2**, its atomic RMSD rising to 10 Å relative to its initial position, by the end of the simulation. During this time, it did not form any any well conserved fibril–ligand interactions (Table 3.3). For LPFFD, the failure of AutoDock to predict a stable pose was even more remarkable. The ligand in position **1**, on the propagation face of the fibril (Figure 3.3a), relocated to the adjacent β-sheet face, with its atomic RMSD rising over 80 Å, relative to its initial position.

Comparing the MM–PBSA predicted binding energies with the ligand sites' AutoDock predicted ranking, three of the MM–PBSA methods used in the analysis above predicted that LHFFD in site **1** had the lowest binding affinity of any ligand–site pair (Table 3.1). The LPFFD ligand that began in site **1** was predicted to form the most favourable fibril–ligand interactions by three of the MM–PBSA methods, but considering the difference between the AutoDock predicted binding site and the final molecular dynamics refined position, it is difficult to describe

AutoDock's prediction as a success. Similarly, all four MM–PBSA methods agree that the LHFFD ligand that began the simulation in site **2**, which AutoDock predicted to be the second most favourable binding site, formed the most stable complex with the Aβ fibril. This ligand completely dissociated from its initial position, before reattaching at a site on the opposite side of the fibril.

It is perhaps unfair to say these results show a failure by AutoDock. We have attempted to perform blind docking of these peptide inhibitors into unknown binding sites, and rank these sites accordingly. This is expecting a lot of both the searching efficiency and scoring accuracy of the docking method. Indeed, molecular docking is more commonly used to dock a large set of molecules into a known target site, and filter out those molecules which are unlikely to be active. The accuracy of a particular docking protocol is usually assessed by its enrichment factor (ie the increase in the percentage of actives in the filtered set compared to the unfiltered set),[88] rather than by the absolute binding energies predicted for a particular ligand. Instead, our results show how molecular dynamics simulation can be used to refine, and assess, the structures of biomolecular complexes obtained from molecular docking, an approach that has been used elsewhere.[89;90]

3.2.3.2 Effects of MM–PBSA Parameters on Calculated Binding Free Energies

The binding free energies of LPFFD and LHFFD in each fibril position were assessed using four MM–PBSA methods (Table 3.1). As there was some variation between the values obtained by these methods, it is necessary to discuss the differences that were found. The most striking difference seen was in the binding energies predicted by method **C**. While the binding free energies obtained via

methods **A**, **B** and **C** agree fairly well, at least qualitatively; other than agreeing on the most favourable binding site for each ligand, method **C**'s energies are not comparable. In fact, two out of three final binding positions were predicted to be unstable by method **C**.

From the binding energy contributions (Table 3.1), the differences in method **C**'s predicted binding energies appear to be largely down to the calculation of the repulsive nonpolar contributions. These contributions to the solvation energy are caused by a combination of two separate physical interactions: firstly the van der Waals repulsion between the atoms of the solute and solvent, and secondly the energy penalty caused by creating a cavity in the solvent. The second interactions usually dominate this term. It has been shown that these hydrophobic interactions vary with solute volume at small sizes (< 10 Å) and solute surface area at larger sizes.[91;92] Method **C** was the only method employed in this work that used a volume dependent repulsive term. As the Aβ fibril is a large peptide aggregate, with many hydrophobic, solvent exposed residues, it is reasonable to suggest that errors in the calculation of the solvation energy of the fibril and complex by method **C** may account for its anomalous binding energies.

3.2.3.3 Comparison of LPFFD– and LHFFD–Fibril Interactions

We performed molecular dynamics refinement of the AutoDock predicted binding positions to see if we could discriminate an active peptide, LPFFD, from an inactive one, LHFFD. MM–PBSA analysis suggested that LPFFD–fibril interaction are stronger than LHFFD-interactions (Table 3.1). All methods (except method **C**, which appears to give erroneous results) predicted that, at equivalent positions, LPFFD was bound more strongly.

The interactions formed between LPFFD and the fibril were more stable than those formed by LHFFD. During the simulations, the LPFFD ligands in all three positions adopted poses within a 5 Å atomic RMSD of their final positions during the final 7.5 ns of simulation. In contrast, only the LHFFD ligand in position **4** appeared to adopt a stable pose. This was also shown during cluster analysis, where a larger clustering radius was required in order to cluster LHFFD conformations (Table 3.3) compared to LPFFD (Table 3.2). The cluster occupancy of each LPFFD ligand (Figure 3.15) shows that the ligands occupy the same clusters at the end of the final 7.5 ns period of simulation as they do at the start, suggesting the ligand is stable. For LHFFD (Figure 3.16), only the ligand in position **4** shows this. In positions **1** and **2**, the cluster in which the ligand will be found varies with time. The ligand clusters are effectively clustered by ligand translation, rather than change in conformations.

The most stable binding sites of both LPFFD and LHFFD lie on the β-sheet face of the fibril, near a groove that lies perpendicular to the fibril chains, formed by the side chains of the His14 and Lys16 residues. Their bound poses are different however. LPFFD lies along the groove (Figure 3.17); while LHFFD straddles it (Figure 3.18). Similar groove-bound modes have been observed in simulations of complexes of dyes with fibril models using shorter peptides. Examples include Congo red with a dodocamer of a short amyloidogenic octapeptide, GNNQQNY, from yeast prion protein Sup35;[93] and Thioflavin T in association with a fibril comprising sixteen seven-residue peptides, Aβ_{16-22}.[94] Groove-bound modes were found for both Congo red and Thioflavin T to be the most favoured, calculated using MM/GBSA as 5.7 and 3.6 kcal/mol respectively more stable than the next preferred mode.[93;94]

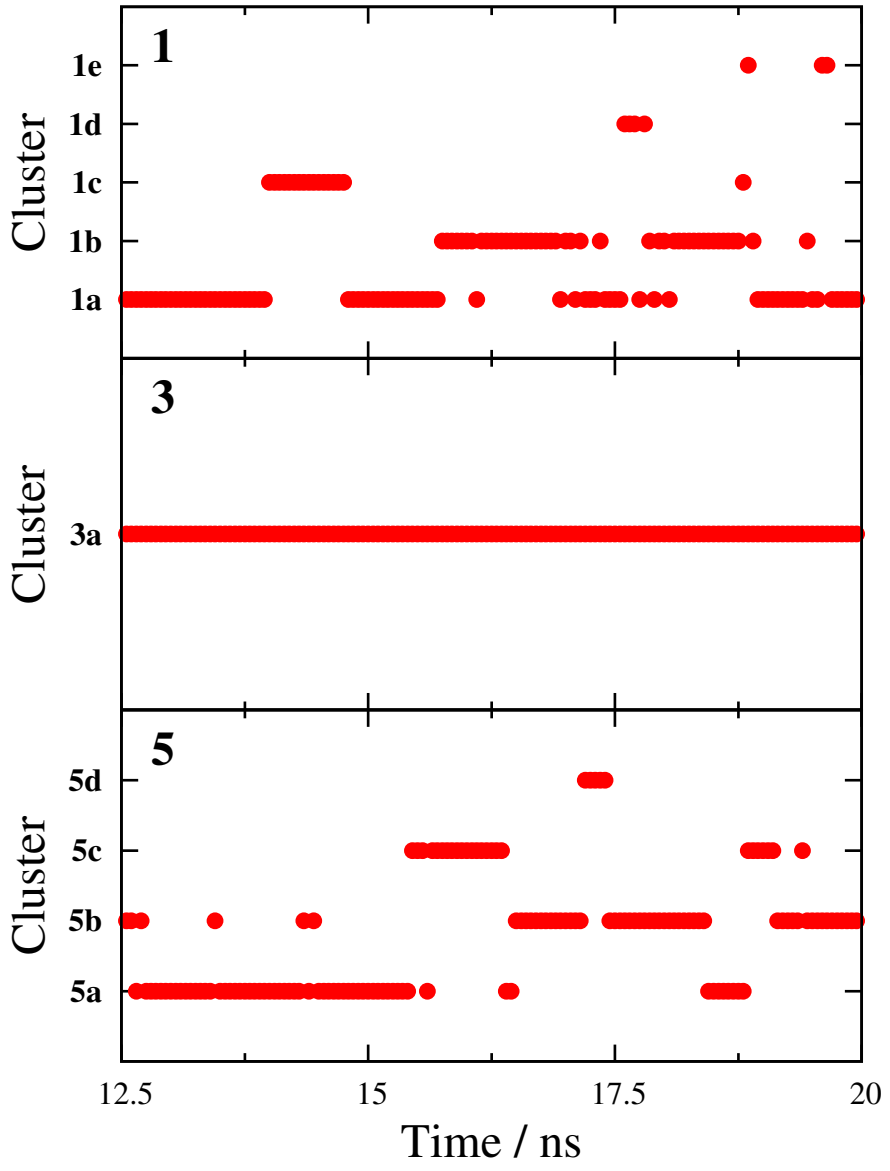


Figure 3.15: Cluster occupancy for LPFFD bound at sites **1**, **3** and **5**.

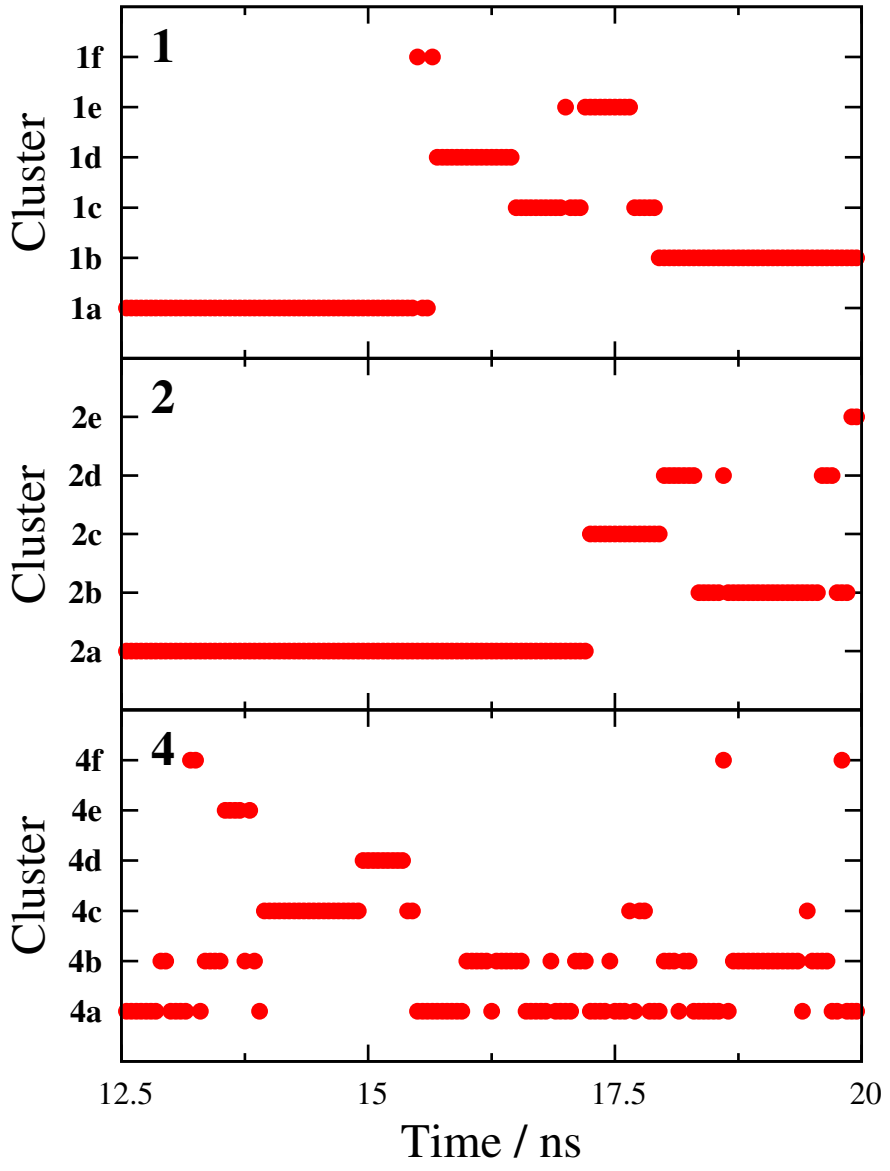
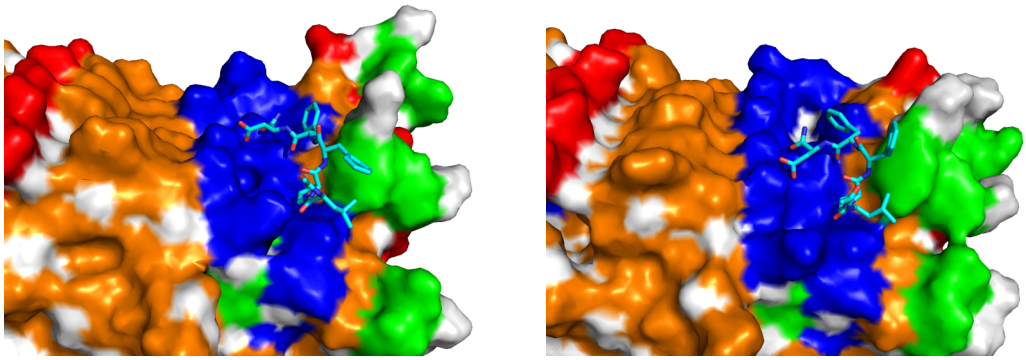
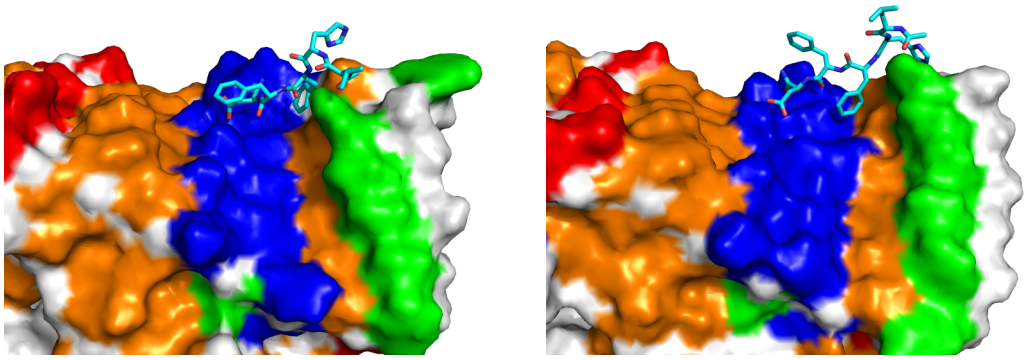


Figure 3.16: Cluster occupancy for LHFFD bound at sites **1**, **2** and **4**.



(a) Cluster **1a**. **(b)** Cluster **1b**.

Figure 3.17: Clusters **1a** and **1b** of LPFFD bound to β-sheet face of fibril (surface colours—red: acidic side chains, blue: basic side chains, green: polar side chains, orange: nonpolar side chains, white: backbone).



(a) Cluster **2a**. **(b)** Cluster **2b**.

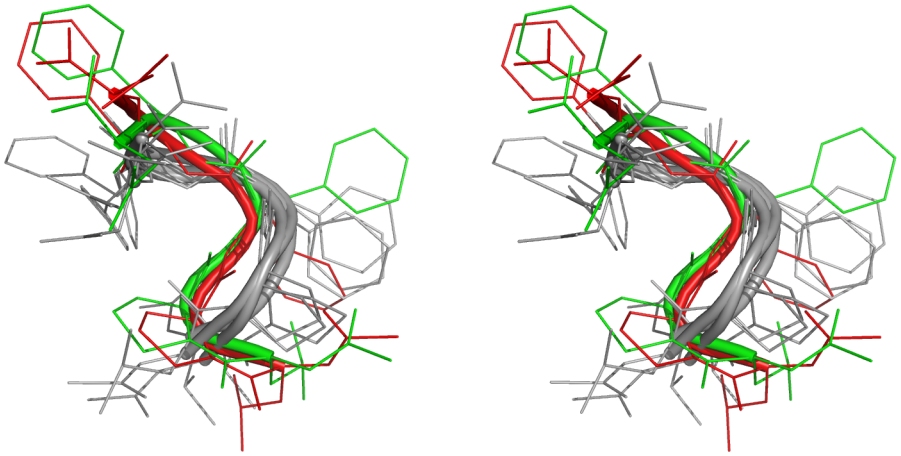
Figure 3.18: Clusters **2a** and **2b** of LHFFD bound to β-sheet face of fibril (surface colours—red: acidic side chains, blue: basic side chains, green: polar side chains, orange: nonpolar side chains, white: backbone).

3.2.3.4 The Role of Proline

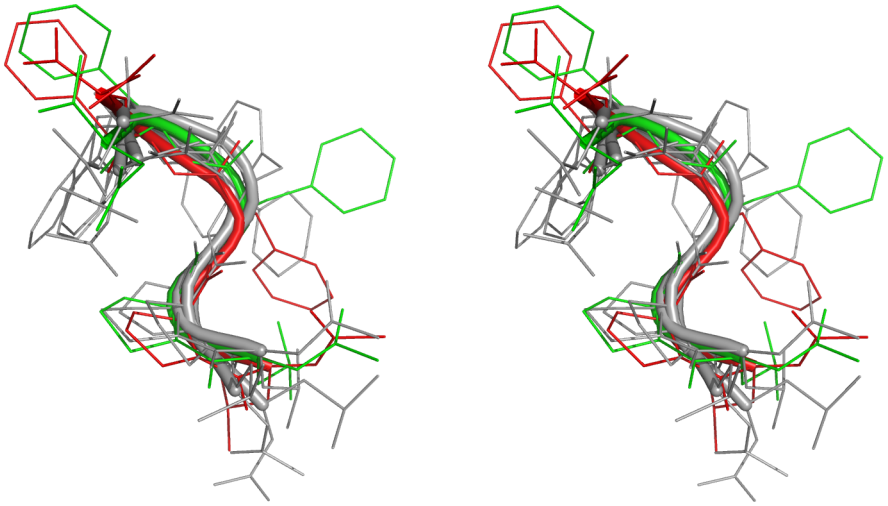
As LPFFD and LHFFD differ by only the substitution of proline for histidine, an important question to ask is what role proline plays in the activity of LPFFD. LPFFD was designed to be complementary to the L_{17}VFF region of the β-sheet structure present in fibrils. Proline was added because of its β-sheet blocking properties, due to its constrained backbone conformation.[70] This work did not find that LPFFD had any strong affinity for the L_{17}VFF region. Instead we found that when bound to the β-sheet face, it took up a position between the His14 and Lys16 residues, as metioned above and shown in Figure 3.17. It may be that the conformationally locked proline in LPFFD is able to disrupt the fibril structure in this position.

The replica exchange simulations of free LPFFD and LHFFD displayed an interesting result. In the simulation of LHFFD, a single, tight turn, peptide backbone conformation was observed that accounted for 86.4% of simulation snapshots (Figure 3.14). While this conformation was the dominant conformation displayed by LPFFD too, the presence of the proline residue in the peptide appears to destabilise it, leading to a second, more open, stable conformation (Figure 3.13b) which accounts for almost a third of LPFFD simulation snapshots.

Interestingly, this second conformation was seen to be a close match to the pose formed by LPFFD, in its top two clusters, when bound to the fibril on the β-sheet face (Figure 3.19b, Table 3.6) or within the hydrophobic lateral cavity (Table 3.6). In comparison, the β-sheet bound pose of LHFFD is a poor match for its conformation in the free state (Figure 3.20, Table 3.7) and it is not until the 8th most populous free cluster, which only accounts for 1.7% of simulation snapshots (Table 3.5), that there is an agreement between the free cluster centroid and



(a) LPFFD clusters **1**, **4** and **6**.



(b) LPFFD clusters **2**, **3** and **5**.

Figure 3.19: Stereoscopic view of six most popular clusters of LPFFD (grey), obtained from replica exchange simulation of free peptide, compared with bound LPFFD poses **1a** (red) and **1b** (green).

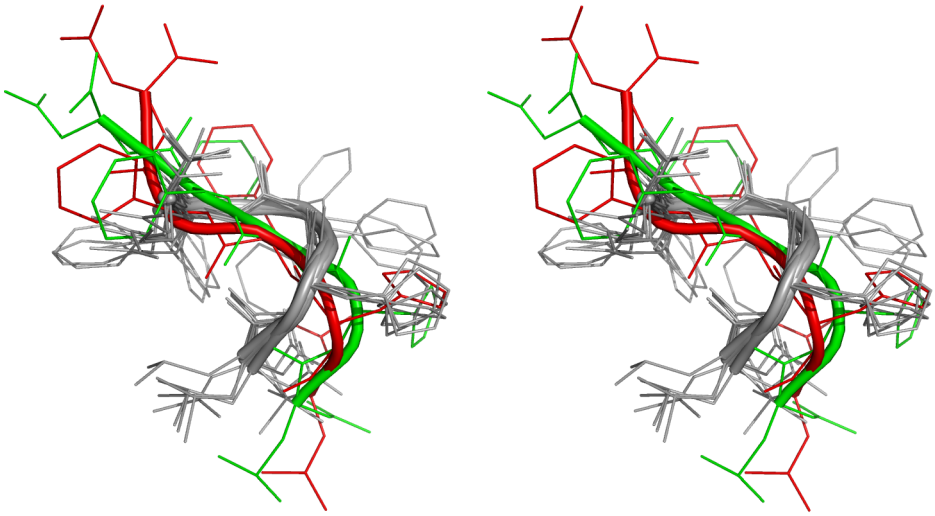


Figure 3.20: Steroscopic view of five most populated clusters ((1) – **5** of LHFFD (grey), obtained freom replica exchange simulation of free peptide, compared with bound LHFFDposes **2a** (red) and **2b** (green).

any of the bound LHFFD clusters (Table 3.7), of less than 1.5 Å RMSD. The complementary shape of LPFFD in both the bound and unbound state, may, in part, explain its higher affinity for fibril binding.

3.3 Aβ Fibril Structure and Free Energy Analysis

In this section we continue the theme of exploring the structure of Aβ fibrils as a target receptor through the use of molecular dynamics simulations and free energy analysis to characterise the mechanical stability of the fibril in the absence of ligands. We employ computational alanine scanning to determine any hot spots in the fibril structure that may be used as potential points of attack by inhibitory ligands.

3.3.1 Methods

3.3.1.1 Choice of Residue Protonation States

As a single monomer of $A\beta_{9-40}$ contains seven titratable residues (Figure 3.21), there are a total of 84 such residues present in the 12-mer structure used in this work. While molecular dynamics methods that allow protonation states to vary during a simulation do exist, it is common practice to assign protonation states to each residue before beginning a simulation and keep these fixed throughout the simulation trajectory. The fibril contains rows of histidine residues. From inspection of their local hydrogen bonding environment, these histidines were modelled with the Nϵ protonated. However in the fibril, there are also two rows of glutamates (Glu11 and Glu22) and a row of lysine residues (Lys16). For these residues, there was not a clear salt bridge partner, and due to their proximity to their corresponding residues in neighbouring peptides, it was decided to model three alternative (or possibly co-existing) protonation states of the assembled fibril. These are denoted: Protonation State 1 – Glu11, Glu22 and Lys28 side-chains

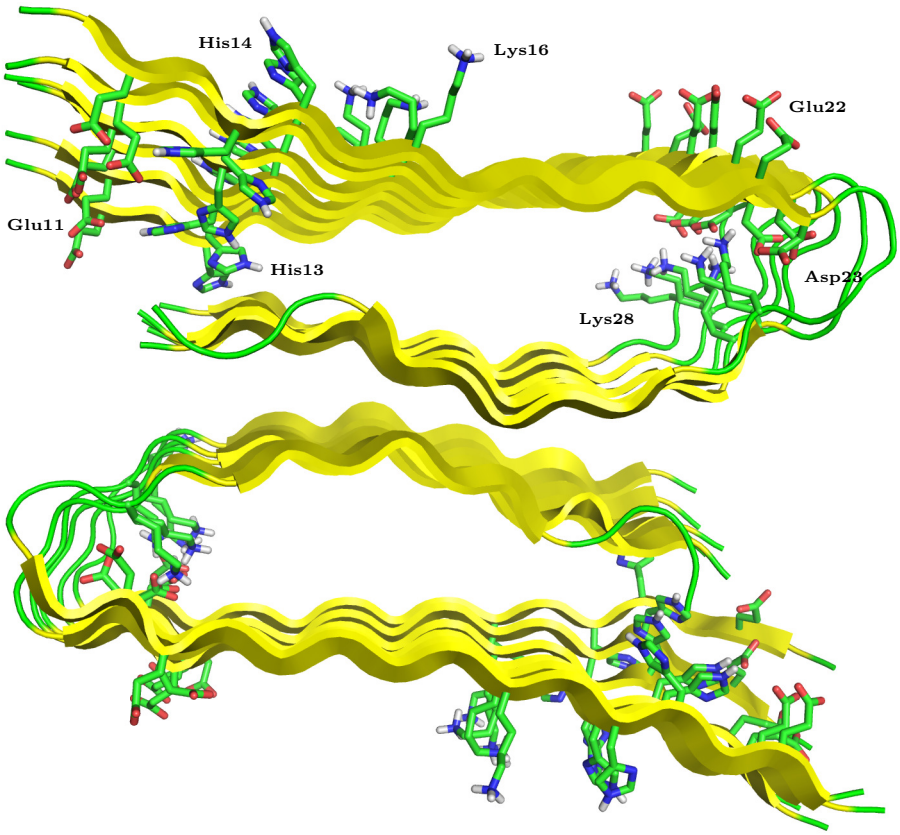


Figure 3.21: Titratable amino acid residues of present in Aβ (residue labelling only shown for one protofibril).

are treated as charged; Protonation State 2 – Glu11 and Glu22 are modelled with neutral, protonated side-chains; Lys28 is charged; Protonation State 3 – Glu11, Glu22 and Lys28 side-chains are treated as uncharged.

3.3.1.2 Simulation of Aβ Fibrils

The three simulations of Aβ fibrils were performed in an identical manner, using the Amber *ff03* force field of Duan et al.[77] The Aβ fibril structure was generated using the *tleap* module of Amber 9,[76] then geometry optimised in *sander* through 500 steps of steepest descent and 500 steps of conjugate gradient minimisation.

During minimisation, the solvent environment was represented by the generalised Born model of Hawkins, Cramer and Truhlar[14;17;18] and non-bonded interactions were not truncated at any distance. The fibril was then solvated in a TIP3P[78] water box using *tleap*.[76] The simulation box of the fibril in protonation state **1** was neutralised by the addition of 12 sodium ions; while the simulation of the fibril in protonation state **2** was neutralised by 12 chlorine ions. In protonation state **3**, the fibril is already neutral. The solvent was then optimised in *sander*[76] by 500 steps of steepest descent and 1000 steps of conjugate gradient minimisation, with harmonic restraints of 500 kcal/mol.Å^2 on the fibril atoms. A non-bonded interaction cut off of 9.0 Å was used, with long range electrostatic represented using particle mesh Ewald.[24] Constraints were placed on all bonds involving hydrogen using the SHAKE algorithm.[79] The whole simulation box was then optimised via a further 500 steps of steepest descent and 1000 steps of conjugate gradient minimisation.

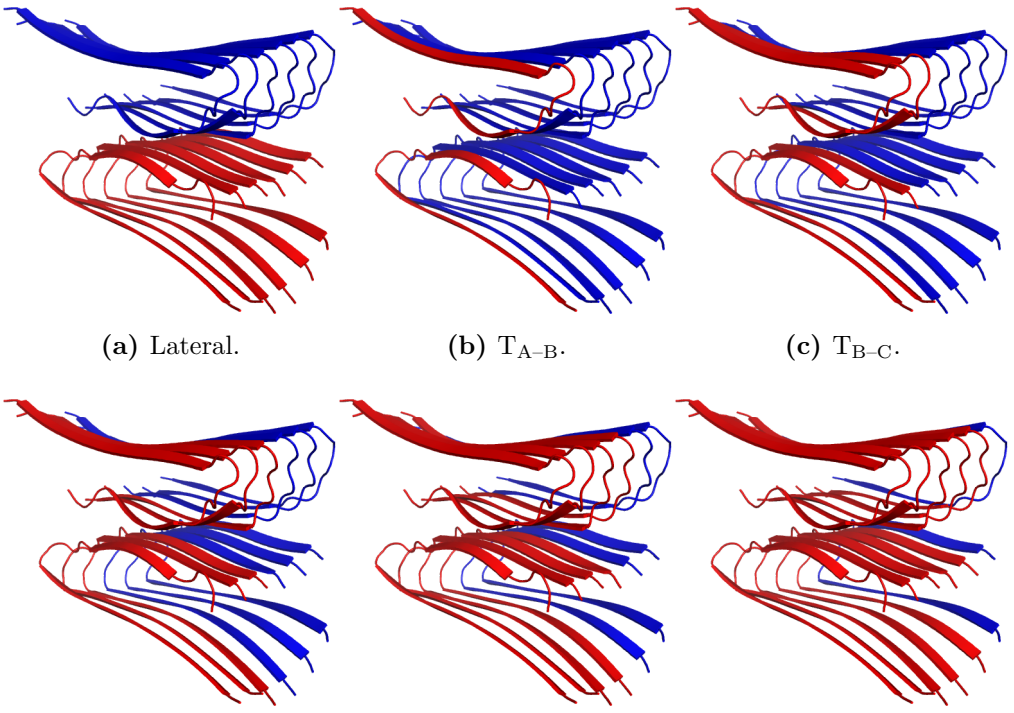
All equilibration and production molecular dynamics simulations were performed using the *pmemd* module of Amber 9,[76] with a simulation time step of 2.0 fs. Long range electrostatic interactions were modelled using the particle mesh Ewald method[24] and non-bonded interactions were truncated at 9 Å. SHAKE[79] was applied to all bonds involving hydrogen. Initial atomic velocities were generated from a Maxwell-Boltzmann distribution at 50 K, then the simulation was heated to 300 K, with harmonic restraints of 5.0 kcal/mol.Å^2 applied to the fibril backbone atoms, over 100 ps of constant volume dynamics. The simulation temperature was controlled through Langevin dynamics,[27] with a collision frequency of 5.0 ps^{-1}. The restraints were then removed and the temperature of the system allowed to equilibrate over 200 ps at constant volume. The density of the simulation box was equilibrated through constant pressure simulation with a weak coupling

barostat.[26] Firstly during a 300 ps simulation using Langevin dynamics[27] with a collision frequency of 5.0 ps^{-1} and a pressure coupling time constant of 2.0 ps; then through a 300 ps simulation using Berendsen's weak coupling thermostat,[26] with a temperature time constant of 1.0 ps and a pressure coupling constant of 2.0 ps. A final 100 ps period of constant volume equilibration was then performed using a temperature coupling constant of 3.0 ps. The 100 ns production simulations proceeded at 300 K, with a coupling time of 5.0 ps, in a constant volume simulation box.

3.3.1.3 MM–PBSA Analysis of Fibril Interfaces

MM–PBSA[43] analysis was used to calculate the interfacial binding free energy of the peptide chain interfaces present in the 12-mer Aβ fibril, for each of the fibril protonation states which were simulated. The binding energies of six interfaces were calculated (Figure 3.22): the lateral interface that lies parallel to the fibril propagation axis and separates protofibrils **I** and **II**, and five transverse interfaces that lie perpendicular to the lateral interface. These transverse interfaces are identified by the fibril chains that they separate, namely between chains **A** and **B** ($T_{A–B}$), between chains **B** and **C** ($T_{B–C}$), between chains **C** and **D** ($T_{C–D}$), between chains **D** and **E** ($T_{D–E}$) and between chains **E** and **F** ($T_{E–F}$).

The interfacial binding energies were calculated using three MM–PBSA methods, **A**, **C** and **D** (see Section 3.2.1.4 for the details of these methods), using *pbsa* from Amber 10.[80] The analysed simulation snapshots were taken at 100 ps intervals throughout the 100 ns simulations. A single trajectory approach was employed, with snapshots of the bound and unbound species taken from a simulation of the 12-mer Aβ fibril.



(a) Lateral. **(b)** T_{A-B}. **(c)** T_{B-C}.

(d) T_{C-D}. **(e)** T_{D-E}. **(f)** T_{E-F}.

Figure 3.22: Interfaces present in 12-mer Aβ fibril. The red and blue coloured chains show the two smaller fibril fragments that would be formed by bisecting the fibril along each interface

Entropic contributions to the binding free energy of each interface was estimated through normal mode analysis,[45;46] using *nab* in Amber 10.[80] Bound and unbound simulation snapshots came from a single trajectory, at an equal spacing of 2 ns. Each snapshot was geometry optimised, with a distance dependent dielectric to represent shielding of charge by the solvent, using conjugate gradient minimisation until the RMS gradient of their molecular mechanics potentials was below 10^{-3} kcal/mol.Å. These optimised geometries were then further minimised by Newton-Raphson optimisation, until their RMS gradients were below 10^{-12} kcal/mol.Å. For some snapshots, the Newton-Raphson procedure was unable to converge to below 10^{-12} kcal/mol.Å following conjugate gradient minimisation. In these cases, further conjugate gradient minimisation was performed until the

RMS gradient of the potentials was below 10^{-4} kcal/mol.Å, before switching to Newton-Raphson. Once the snapshot geometries were optimised, they were used for normal mode calculations of the bound and unbound species, in a distance dependent dielectric.

3.3.1.4 Computational Alanine Scanning

The contribution of each fibril residue to the binding free energy T_{CD} interface was assessed using MM–PBSA computational alanine scanning.[95] For each residue that makes up the T_{CD} interface, a full set of simulation snapshots was generated in which the residue of interest is mutated to alanine. This mutation was performed using the *mm_pbsa.pl* script of Amber 9,[76] which deletes each atom, other than the C_α atom, bound to the C_β atom of the residue, and places a hydrogen 1.09 Å from C_β, in the C_β–deleted atom bond direction. MM–PBSA analysis was then performed using the set of singly mutated snapshots. The difference between the interfacial binding free energy calculated using the mutated snapshots and the wild-type snapshots ($\Delta\Delta G_{bind}$) was used as an indication of the the importance of that residue to the binding strength of the interface (Equation 3.2).

$$\Delta\Delta G_{bind} = \Delta G_{bind}^{mutant} - \Delta G_{bind}^{WT} \tag{3.2}$$

As four Aβ monomers make up the T_{CD} interface, chains **C** and **D** of protofibrils **I** and **II**, the error in the calculation of $\Delta\Delta G_{bind}$ for each residue was estimated from the standard deviation in the calculated value for each monomer.

3.3.1.5 Computational Resources

Each of the three 100 ns fibril simulations took around 500 hours to run on 64 processor cores. MM-PBSA analysis of each interface required almost 50 CPU hours per MM-PBSA method, plus around 1500 hours for the entropy calculations. Calculation of the interfacial energies of the three fibril ionisation state, by three MM-PBSA methods, therefore required a total of over 10000 CPU hours. Similarly, each alanine scanning mutation required approximately 30 hours, per MM-PBSA method (less than the interface calculations as some of the necessary calculations had already been performed). 96 scanning mutations were performed, using 3 MM-PBSA methods, requiring a total CPU time of approximately 8500 hours.

3.3.2 Results

3.3.2.1 Fibril Structural Analysis

During the simulation of the Aβ fibril in protonation state **1**, the fibril appeared to have relaxed into a stable state within 20 ns, with its backbone atom RMSD settling to a value below 4 Å, relative to the initial simulation coordinates (Figure 3.23a). After 50 ns, instability in the fibril structure resulted in a disruption of its structure, with the backbone atomic RMSD rising to ~6 Å by the end of the 100 ns simulation.

The fibril in protonation state **2** appeared to be more stable during the simulation (Figure 3.23b). During the final 70 ns the backbone RMSD of the fibril remained approximately 3 Å from the starting fibril geometry. During the simulation of the

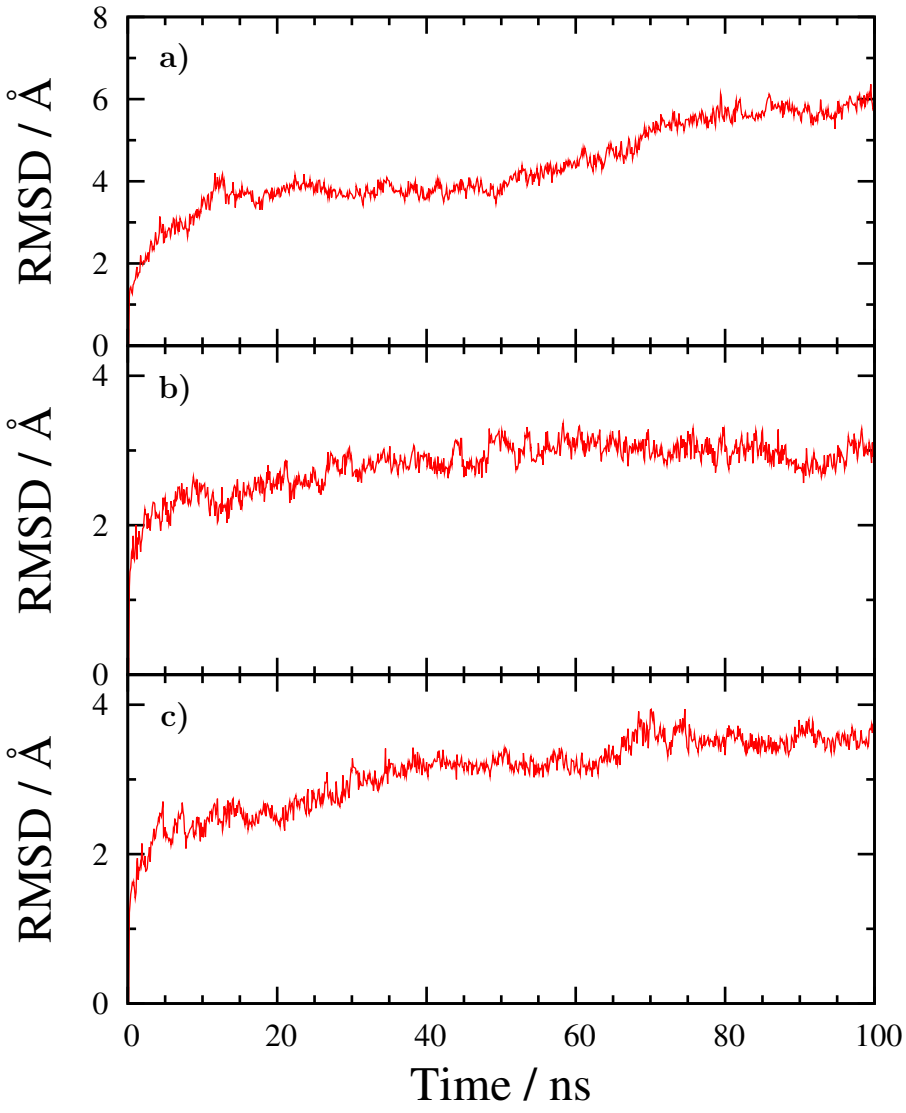


Figure 3.23: Backbone RMSD (Å) of Aβ fibril during 100 ns simulation, for a) protonation **1**, b) protonation state **2** and c) protonation state **3**.

fibril in protonation state **3**, the backbone RMSD rose steadily then stabilised at around 3.2 Å between 35 and 65 ns. Following this period, the RMSD again rose, before stabilising at approximately 3.6 Å.

3.3.2.2 Interfacial Binding Free Energies

In order to assess the relative strengths of the fibril interfaces in each protonation state, MM–PBSA analysis was performed (Tables 3.8 to 3.10). For the simulation of the fibril in protonation state **1**, equally space snapshots were taken from the stable period between 20 and 50 ns, giving 300 snapshots for the calculation of the molecular mechanics and solvation components, and 15 snapshots for the entropy calculations.

Methods **A** and **D** predicted that the lateral interface, between the two protofibril horseshoes, was the weakest interface in the fibril, with total binding free energies of −63.4 and −93.4 kcal/mol respectively (Table 3.8). The binding affinities of the transverse interfaces were predicted to be 2–3 times higher than the lateral interface. For example, for method **A**, the transverse interfacial energies ranged from −238.1 to −303.2 kcal/mol. There did not appear to be any trend relating to the position of the interface within the fibril, ie. edge relative to the middle of the stack. The narrow range in total binding free energies is mirrored by the binding free energies in the absence of the solute entropic correction (Table 3.8). The van der Waals contributions are more broadly distributed: for the transverse interfaces, this is found to be −275.5 to −389.0 kcal/mol (Table 3.8). This compares with a van der Waals binding contribution of −202.6 kcal/mol for the lateral interface. There is larger variation in the polar solvation contributions across the transverse interfaces, with estimates ranging from +127.8 kcal/mol to

−369.7 kcal/mol (Table 3.8). This appears to reflect a considerable sensitivity of polar solvation energy estimate to the orientation of charged residues at the interface. These trends are reflected in the polar solvation contributions estimated by method **D** (and therefore also by method **C**).

The interfacial binding affinities predicted by method **C** differed greatly from those given by the other two methods, both in magnitude and in ranking order. Indeed this method gave positive binding free energies for all interfaces in the fibril (Table 3.8). The differences between method **C** and methods **A** and **D** were largely in the non-polar parts of the calculation, with Method **C** giving an overall unfavourable non-polar interaction between each unbound species pair. All interfaces, except the $T_{A–B}$ interface, showed a positive electrostatic contribution to their binding free energy and a negative Poisson-Boltzmann one.

MM–PBSA analysis was performed using snapshots from the final 70 ns of the simulation of the fibril in protonation state **2** where residues Glu11 and Glu22 are protonated. This gave 700 snapshots over which to average the Poisson-Boltzmann and molecular mechanics contributions, and 35 snapshots for the entropy calculations.

The lateral interface was again predicted to be the weakest interface of the fibril by methods **A** and **D** (Table 3.9). As before, for both models, binding of fibril across the lateral interface appears to be approximately half as high in affinity as across any of the transverse interfaces. Method **C** predicted that the lateral interface would have the least favourable interactions due to the molecular mechanics and solvation terms. However, as binding at this interface resulted in the smallest entropic penalty, the $T_{B–C}$, $T_{C–D}$ and $T_{D–E}$ interfaces had less favourable total binding free energies, 56.7 – 69.9 kcal/mol, compared to 52.0 kcal/mol for the

Table 3.8: Binding free energy (kcal/mol) of Aβ fibril interfaces, for protonation state **1**, with standard deviations. Calculations performed using three different PBSA methods (see section 3.2.1.4).

Ligand	[a] ΔG_{bind}^{total}	[b] $\Delta G_{bind}^{MM-PBSA}$	[c] ΔG_{elec}	[d] ΔG_{vdW}	[e] ΔG_{PB}	[f] ΔG_{rep}	[g] ΔG_{dis}	[h] ΔG_{np}	[i] $-T\Delta S$
Method A									
Lateral	−63.4 ± 20.4	−136.8 ± 10.6	209.8 ± 57.4	−202.6 ± 10.5	−116.2 ± 57.6	na	na	−27.8 ± 1.5	73.4 ± 17.4
$T_{A–B}$	−147.1 ± 25.2	−238.1 ± 15.7	-52.6 ± 70.0	−275.5 ± 11.4	127.8 ± 64.6	na	na	−37.8 ± 1.5	91.0 ± 19.7
$T_{B–C}$	−143.7 ± 31.8	−251.1 ± 21.5	206.3 ± 57.2	−330.9 ± 11.8	−84.2 ± 54.9	na	na	−42.4 ± 1.0	107.4 ± 23.4
$T_{C–D}$	−165.2 ± 23.5	−273.2 ± 20.0	525.1 ± 83.3	−380.8 ± 11.3	−369.7 ± 72.3	na	na	−47.9 ± 0.8	108.0 ± 12.3
$T_{D–E}$	−196.2 ± 27.5	−303.2 ± 20.7	474.7 ± 118.5	−389.8 ± 13.1	−339.6 ± 100.3	na	na	−48.5 ± 1.0	107.0 ± 18.1
$T_{E–F}$	−188.3 ± 25.9	−270.7 ± 16.1	256.5 ± 95.7	−325.0 ± 10.8	−158.3 ± 81.0	na	na	−43.9 ± 1.3	82.4 ± 20.3
Method C									
Lateral	95.2 ± 22.0	21.8 ± 13.4	209.8 ± 57.4	−202.6 ± 10.5	−113.4 ± 60.7	−137.8 ± 7.4	265.9 ± 11.5	128.1 ± 4.9	73.4 ± 17.4
$T_{A–B}$	61.5 ± 28.5	−29.5 ± 20.6	-52.6 ± 70.0	−275.5 ± 11.4	141.0 ± 64.0	−196.5 ± 8.2	354.1 ± 11.7	157.6 ± 4.9	91.0 ± 19.7
$T_{B–C}$	134.7 ± 34.9	27.3 ± 25.9	206.3 ± 57.2	−330.9 ± 11.8	−51.4 ± 53.1	−220.7 ± 6.4	424.0 ± 8.6	203.3 ± 3.7	107.4 ± 23.4
$T_{C–D}$	136.5 ± 24.6	28.5 ± 21.3	525.1 ± 83.3	−380.8 ± 11.3	−343.7 ± 74.4	−254.2 ± 5.3	482.1 ± 7.2	227.9 ± 3.6	108.0 ± 12.3
$T_{D–E}$	92.2 ± 31.0	−14.8 ± 25.1	474.7 ± 118.5	−389.8 ± 13.1	−322.0 ± 103.1	−260.0 ± 6.4	482.5 ± 8.6	222.5 ± 3.9	107.0 ± 18.1
$T_{E–F}$	54.3 ± 29.7	−28.1 ± 21.7	256.5 ± 95.7	−325.0 ± 10.8	−141.5 ± 82.4	−227.9 ± 6.4	409.7 ± 10.2	181.9 ± 5.0	82.4 ± 20.3
Method D									
Lateral	−93.4 ± 23.9	−166.8 ± 16.3	209.8 ± 57.4	−202.6 ± 10.5	−113.4 ± 60.7	−326.4 ± 17.6	265.9 ± 11.5	−60.6 ± 7.5	73.4 ± 17.4
$T_{A–B}$	−207.3 ± 31.7	−298.3 ± 24.9	-52.6 ± 70.0	−275.5 ± 11.4	141.0 ± 64.0	−465.3 ± 19.4	354.1 ± 11.7	−111.2 ± 9.3	91.0 ± 19.7
$T_{B–C}$	−167.0 ± 38.0	−274.4 ± 30.0	206.3 ± 57.2	−330.9 ± 11.8	−51.4 ± 53.1	−522.5 ± 15.1	424.0 ± 8.6	−98.5 ± 7.9	107.4 ± 23.4
$T_{C–D}$	−210.9 ± 26.8	−318.9 ± 23.8	525.1 ± 83.3	−380.8 ± 11.3	−343.7 ± 74.4	−601.6 ± 12.6	482.1 ± 7.2	−119.5 ± 7.2	108.0 ± 12.3
$T_{D–E}$	−263.2 ± 31.9	−370.2 ± 26.2	474.7 ± 118.5	−389.8 ± 13.1	−322.0 ± 103.1	−615.5 ± 15.1	482.5 ± 8.6	−133.0 ± 8.2	107.0 ± 18.1
$T_{E–F}$	−257.2 ± 29.4	−339.6 ± 21.2	256.5 ± 95.7	−325.0 ± 10.8	−141.5 ± 82.4	−539.4 ± 15.2	409.7 ± 10.2	−129.6 ± 7.2	82.4 ± 20.3

[a] Total MM–PBSA binding free energy.
[b] MM–PBSA binding free energy, excluding entropic contributions.
[c] Electrostatic contribution to the binding free energy.
[d] van der Waals contribution to the binding free energy.
[e] Poisson-Boltzmann electrostatic solvation contribution to the binding free energy.
[f] Non-polar repulsive contribution to the binding free energy.
[g] Non-polar dispersion contribution to the binding free energy.
[h] Total non-polar contribution to the binding free energy.
[i] Entropic contribution to the binding free energy.

lateral interface (Table 3.9). Method **C** predicted positive total binding free energies for all but the T_{EF} interface.

It appears that neutralisation of the charged Glu11 and Glu22 residues removes much of the electrostatic strain between the fibril interfaces, with all but the $T_{C–D}$ interface showing favourable electrostatic contributions to their binding free energy. As a result of this, all MM–PBSA methods predict an increase in the binding affinity of all interfaces, compared to protonation state **1** (Tables 3.8 and 3.9). For example, according to method **A**, the total binding free energy for the lateral interface is predicted as −113.9 kcal/mol for protonated glutamates (Table 3.25) as compared with −63.4 kcal/mol for deprotonated glutamates (Table 3.24). As for protonation state **1**, there is considerable variation in the polar solvation energy contribution to binding (eg. for model A, 60.2 – 245.7 kcal/mol, Table 3.9). Nevertheless, this is a smaller range than for protonation state 1 with its more highly charged interfaces. Also, the van der Waals contribution appears to span a narrower range across transverse interfaces (of 59.1 kcal/mol Table 3.9), relative to fibril in protonation state **1**. This may accord with the lower RMSD found in the protonation state **2** simulation (Figure 3.23).

Finally, for a fibril with neutral Glu11, Glu22 and Lys16 residues (protonation state **3**), structures from 35 to 65 ns into the 100 ns simulation were used for MM–PBSA analysis of the interfacial binding free energies (Table 3.10). The molecular mechanics and solvation terms in the calculation were averaged over 300 equally spaced snapshots from this time period, while 15 snapshots were used for the entropy calculation.

As in the previous two cases, methods **A** and **D** suggested that the lateral interface was the weakest of the fibril interfaces, by a factor of around 2-3 relative

Table 3.9: Binding free energy (kcal/mol) of Aβ fibril interfaces, for protonation state **2**, with standard deviations. Calculations performed using three different PBSA methods (see section 3.2.1.4).

Ligand	[a] ΔG_{bind}^{total}	[b] $\Delta G_{bind}^{MM-PBSA}$	[c] ΔG_{elec}	[d] ΔG_{vdW}	[e] ΔG_{PB}	[f] ΔG_{rep}	[g] ΔG_{dis}	[h] ΔG_{np}	[i] $-T\Delta S$
Method A									
Lateral	−113.9 ± 17.9	−171.5 ± 9.1	−100.6 ± 49.0	−228.0 ± 8.5	185.7 ± 47.3	na	na	−28.6 ± 0.8	57.6 ± 15.5
$T_{A–B}$	−231.7 ± 23.3	−316.2 ± 13.4	−149.6 ± 51.0	−333.8 ± 12.6	210.1 ± 51.2	na	na	−42.9 ± 1.6	84.5 ± 19.0
$T_{B–C}$	−252.7 ± 22.7	−359.1 ± 13.8	−104.8 ± 66.4	−381.4 ± 10.7	173.6 ± 67.7	na	na	−46.5 ± 0.8	106.4 ± 18.0
$T_{C–D}$	−224.3 ± 22.6	−340.4 ± 15.8	40.3 ± 64.8	−393.7 ± 13.1	60.2 ± 59.4	na	na	−47.2 ± 0.9	116.1 ± 16.1
$T_{D–E}$	−217.2 ± 21.2	−321.7 ± 15.9	−25.6 ± 65.7	−360.7 ± 11.7	111.3 ± 58.8	na	na	−46.7 ± 1.1	104.5 ± 14.1
$T_{E–F}$	−286.6 ± 26.4	−363.6 ± 14.8	−204.5 ± 61.9	−356.0 ± 10.5	245.7 ± 56.4	na	na	−48.8 ± 0.8	77.0 ± 21.9
Method C									
Lateral	52.9 ± 20.8	−4.7 ± 13.9	−100.6 ± 49.0	−228.0 ± 8.5	187.6 ± 49.4	−149.1 ± 4.7	285.4 ± 6.6	136.3 ± 3.5	57.6 ± 15.5
$T_{A–B}$	5.9 ± 26.8	−78.6 ± 18.8	−149.6 ± 51.0	−333.8 ± 12.6	220.9 ± 51.7	−230.2 ± 8.3	414.1 ± 13.4	183.9 ± 6.2	84.5 ± 19.0
$T_{B–C}$	56.7 ± 25.3	−49.7 ± 17.7	−104.8 ± 66.4	−381.4 ± 10.7	209.7 ± 66.0	−252.0 ± 5.7	478.9 ± 7.4	226.8 ± 3.6	106.4 ± 18.0
$T_{C–D}$	69.9 ± 26.6	−46.2 ± 21.2	40.3 ± 64.8	−393.7 ± 13.1	73.5 ± 59.7	−255.6 ± 5.9	489.2 ± 8.0	233.6 ± 3.8	116.1 ± 16.1
$T_{D–E}$	62.2 ± 24.2	−42.3 ± 19.7	−25.6 ± 65.7	−360.7 ± 11.7	123.9 ± 58.6	−245.8 ± 6.6	465.9 ± 8.4	220.1 ± 3.5	104.5 ± 14.1
$T_{E–F}$	−25.2 ± 28.6	−102.2 ± 18.5	−204.5 ± 61.9	−356.0 ± 10.5	268.3 ± 55.6	−260.1 ± 4.7	450.2 ± 5.9	190.1 ± 3.7	77.0 ± 21.9
Method D									
Lateral	−151.0 ± 22.3	−208.6 ± 16.1	−100.6 ± 49.0	−228.0 ± 8.5	187.6 ± 49.4	−353.0 ± 11.1	285.4 ± 6.6	−67.6 ± 6.3	57.6 ± 15.5
$T_{A–B}$	−308.7 ± 28.8	−393.2 ± 21.6	−149.6 ± 51.0	−333.8 ± 12.6	220.9 ± 51.7	−544.8 ± 19.7	414.1 ± 13.4	−130.7 ± 8.4	84.5 ± 19.0
$T_{B–C}$	−287.8 ± 27.8	−394.2 ± 21.2	−104.8 ± 66.4	−381.4 ± 10.7	209.7 ± 66.0	−596.5 ± 13.6	478.9 ± 7.4	−117.7 ± 8.0	106.4 ± 18.0
$T_{C–D}$	−279.4 ± 29.2	−395.5 ± 24.3	40.3 ± 64.8	−393.7 ± 13.1	73.5 ± 59.7	−604.9 ± 14.1	489.2 ± 8.0	−115.7 ± 7.8	116.1 ± 16.1
$T_{D–E}$	−273.7 ± 27.6	−378.2 ± 23.7	−25.6 ± 65.7	−360.7 ± 11.7	123.9 ± 58.6	−581.7 ± 15.5	465.9 ± 8.4	−115.8 ± 8.5	104.5 ± 14.1
$T_{E–F}$	−380.7 ± 30.5	−457.7 ± 21.3	−204.5 ± 61.9	−356.0 ± 10.5	268.3 ± 55.6	−615.6 ± 11.2	450.2 ± 5.9	−165.4 ± 7.6	77.0 ± 21.9

[a] Total MM–PBSA binding free energy.
[b] MM–PBSA binding free energy, excluding entropic contributions.
[c] Electrostatic contribution to the binding free energy.
[d] van der Waals contribution to the binding free energy.
[e] Poisson-Boltzmann electrostatic solvation contribution to the binding free energy.
[f] Non-polar repulsive contribution to the binding free energy.
[g] Non-polar dispersion contribution to the binding free energy.
[h] Total non-polar contribution to the binding free energy.
[i] Entropic contribution to the binding free energy.

to the transverse interfaces (Table 3.10). Method **C** predicted that the lateral interface was weaker before the entropic contributions were accounted for. The neutralisation of the charged Lys16 residue resulted in increasingly favourable electrostatic interactions between the interfaces, presumably due to a decrease in the repulsive interaction between the adjacent charged groups. However, this was accompanied by corresponding decreases in the Poisson-Boltzmann contributions to binding. As for protonation state **2**, the range of these differences is smaller than for the more highly charged interfaces of protonation state **1**.

Table 3.10: Binding free energy (kcal/mol) of Aβ fibril interfaces, for protonation state **3**, with standard deviations. Calculations performed using three different PBSA methods (see section 3.2.1.4).

Ligand	[a] ΔG_{bind}^{total}	[b] $\Delta G_{bind}^{MM-PBSA}$	[c] ΔG_{elec}	[d] ΔG_{vdW}	[e] ΔG_{PB}	[f] ΔG_{rep}	[g] ΔG_{dis}	[h] ΔG_{np}	[i] $-T\Delta S$
Method A									
Lateral	−102.4 ± 19.5	−168.9 ± 14.5	−331.8 ± 60.4	−207.8 ± 7.2	399.3 ± 55.8	na	na	−28.6 ± 0.9	66.5 ± 13.1
$T_{A–B}$	−212.3 ± 20.6	−291.3 ± 12.4	−339.8 ± 42.1	−306.2 ± 9.9	396.8 ± 41.4	na	na	−42.1 ± 1.2	79.0 ± 16.4
$T_{B–C}$	−226.6 ± 26.9	−324.3 ± 13.6	−356.1 ± 43.5	−343.7 ± 10.7	420.4 ± 42.1	na	na	−44.9 ± 1.0	97.7 ± 23.2
$T_{C–D}$	−256.2 ± 23.5	−363.3 ± 12.3	−508.3 ± 58.9	−369.5 ± 12.1	561.1 ± 52.2	na	na	−46.5 ± 1.0	107.1 ± 20.0
$T_{D–E}$	−252.1 ± 22.9	−357.8 ± 16.5	−602.4 ± 86.0	−385.6 ± 13.4	680.0 ± 78.7	na	na	−49.8 ± 1.0	105.7 ± 15.8
$T_{E–F}$	−249.1 ± 29.4	−334.3 ± 15.9	−455.8 ± 56.0	−348.8 ± 11.2	517.4 ± 52.7	na	na	−47.1 ± 0.9	85.2 ± 24.7
Method C									
Lateral	74.6 ± 21.1	8.1 ± 16.6	−331.8 ± 60.4	−207.8 ± 7.2	420.8 ± 60.6	−144.0 ± 4.7	270.9 ± 6.5	126.9 ± 3.4	66.5 ± 13.1
$T_{A–B}$	1.0 ± 23.8	−78.0 ± 17.3	−339.8 ± 42.1	−306.2 ± 9.9	400.3 ± 42.1	−218.6 ± 5.9	386.4 ± 8.1	167.7 ± 3.8	79.0 ± 16.4
$T_{B–C}$	54.1 ± 31.4	−43.6 ± 21.1	−356.1 ± 43.5	−343.7 ± 10.7	446.4 ± 42.3	−233.7 ± 5.5	443.5 ± 8.6	209.8 ± 4.5	97.7 ± 23.2
$T_{C–D}$	38.6 ± 26.9	−68.5 ± 18.0	−508.3 ± 58.9	−369.5 ± 12.1	585.6 ± 51.1	−248.5 ± 5.3	472.2 ± 8.4	223.7 ± 4.6	107.1 ± 20.0
$T_{D–E}$	49.0 ± 26.4	−56.7 ± 21.1	−602.4 ± 86.0	−385.6 ± 13.4	702.3 ± 85.2	−263.3 ± 6.6	492.3 ± 9.4	229.0 ± 4.3	105.7 ± 15.8
$T_{E–F}$	6.1 ± 30.9	−79.1 ± 18.6	−455.8 ± 56.0	−348.8 ± 11.2	534.0 ± 52.7	−247.0 ± 5.8	438.6 ± 7.6	191.5 ± 4.0	85.2 ± 24.7
Method D									
Lateral	−122.4 ± 23.4	−188.9 ± 19.4	−331.8 ± 60.4	−207.8 ± 7.2	420.8 ± 60.6	−341.0 ± 11.2	270.9 ± 6.5	−70.1 ± 6.5	66.5 ± 13.1
$T_{A–B}$	−297.9 ± 26.4	−376.9 ± 20.7	−339.8 ± 42.1	−306.2 ± 9.9	400.3 ± 42.1	−517.5 ± 14.0	386.4 ± 8.1	−131.2 ± 7.7	79.0 ± 16.4
$T_{B–C}$	−265.5 ± 32.5	−363.2 ± 22.8	−356.1 ± 43.5	−343.7 ± 10.7	446.4 ± 42.3	−553.3 ± 13.1	443.5 ± 8.6	−109.7 ± 6.7	97.7 ± 23.2
$T_{C–D}$	−301.1 ± 27.6	−408.2 ± 19.0	−508.3 ± 58.9	−369.5 ± 12.1	585.6 ± 51.1	−588.2 ± 12.4	472.2 ± 8.4	−116.0 ± 6.5	107.1 ± 20.0
$T_{D–E}$	−310.9 ± 28.8	−416.6 ± 24.1	−602.4 ± 86.0	−385.6 ± 13.4	702.3 ± 85.2	−623.2 ± 15.6	492.3 ± 9.4	−130.9 ± 8.1	105.7 ± 15.8
$T_{E–F}$	−331.6 ± 33.9	−416.8 ± 23.3	−455.8 ± 56.0	−348.8 ± 11.2	534.0 ± 52.7	−584.8 ± 13.7	438.6 ± 7.6	−146.2 ± 8.1	85.2 ± 24.7

[a] Total MM–PBSA binding free energy.
[b] MM–PBSA binding free energy, excluding entropic contributions.
[c] Electrostatic contribution to the binding free energy.
[d] van der Waals contribution to the binding free energy.
[e] Poisson-Boltzmann electrostatic solvation contribution to the binding free energy.
[f] Non-polar repulsive contribution to the binding free energy.
[g] Non-polar dispersion contribution to the binding free energy.
[h] Total non-polar contribution to the binding free energy.
[i] Entropic contribution to the binding free energy.

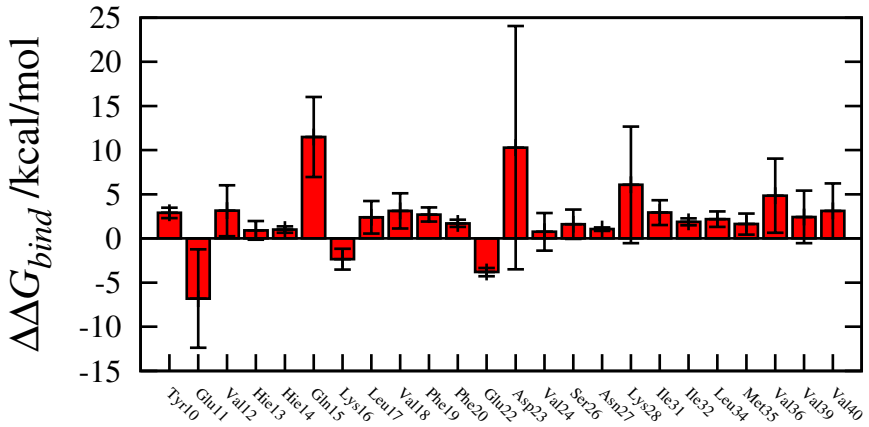
3.3.2.3 Computational Alanine Scanning of Fibril $T_{C–D}$ Interface

Computational alanine scanning of the $T_{C–D}$ interface, found in the middle of the fibril, was performed in order to identify which Aβ residues play an important role in the strength of its binding. All three MM–PBSA methods predicted that the largest decrease in binding affinity occurred on mutation of the Gln15 residue of the fibril in protonation state **1**, with mutation of this residue giving a $\Delta\Delta G_{bind}$ of 11.5 ± 4.5, 14.5 ± 3.9 and 17.4 ± 5.2 kcal/mol, as calculated by methods **A**, **C** and **D** respectively (Figure 3.24). The next highest $\Delta\Delta G_{bind}$ occurred upon mutation of the Asp23 residue, with $\Delta\Delta G_{bind}$ calculated as 10.3 ± 13.8, 12.7 ± 16.8 and 13.6 ± 17.4 kcal/mol by methods **A**, **C** and **D**. The large error bars on this value, as shown in Figure 3.24, should be noted. The reason for this large error is that this residue forms a salt bridge with the Lys28 that crosses the transverse interface. As the direction of this interaction does not change during the simulation, but remains as found in the experimental structure, only two of the four Asp23 residues that are mutated are involved in interactions that cross the $T_{C–D}$ interface (Table 3.11). The importance of this interaction to the stability of the transverse interface is therefore underestimated by the average value shown in Figure 3.24.

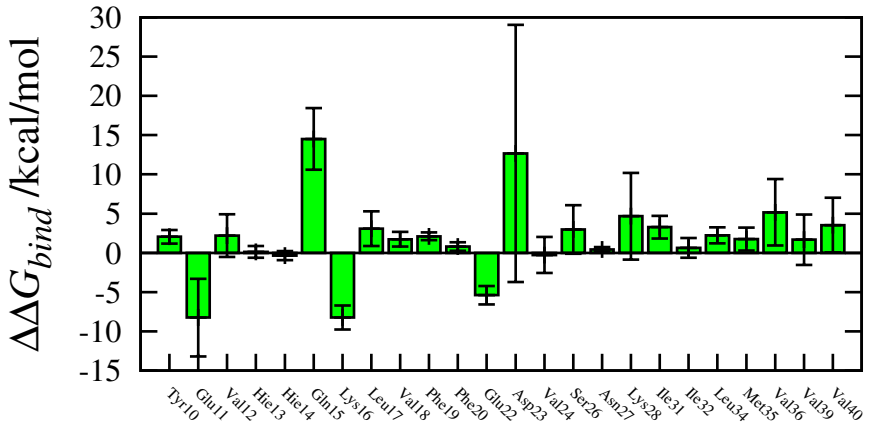
Table 3.11: $\Delta\Delta G_{bind}$ (kcal/mol) for alanine scanning of the Asp23 residue.

Residue	Method **A**	Method **C**	Method **D**
$Asp23_{C(I)}$	−1.9	−1.8	−1.8
$Asp23_{C(II)}$	−1.4	−0.9	−0.9
$Asp23_{D(I)}$	22.0	23.9	25.8
$Asp23_{D(II)}$	22.4	29.5	31.3

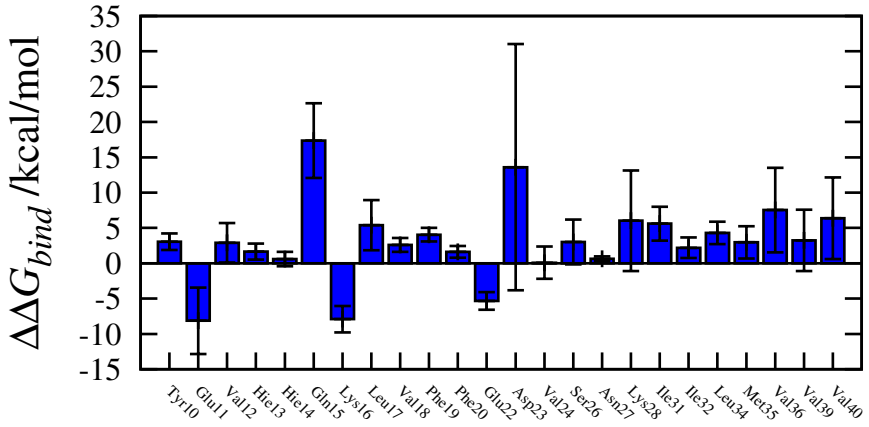
Similarly, the mean $\Delta\Delta G_{bind}$ values calculated for the mutation of the Lys28 residues, 6.0 ± 6.6, 4.6 ± 5.5 and 6.0 ± 7.1 kcal/mol respectively for methods **A**,



(a) Method **A**



(b) Method **C**



(c) Method **D**

Figure 3.24: Computational alanine scanning of the T_{C-D} interface of the fibril in protonation state **1**. Error bars show standard deviation across four fibril monomers.

C and **D**, also underestimate the importance of this residue (Table 3.12).

Table 3.12: $\Delta\Delta G_{bind}$ (kcal/mol) for alanine scanning of the Lys28 residue.

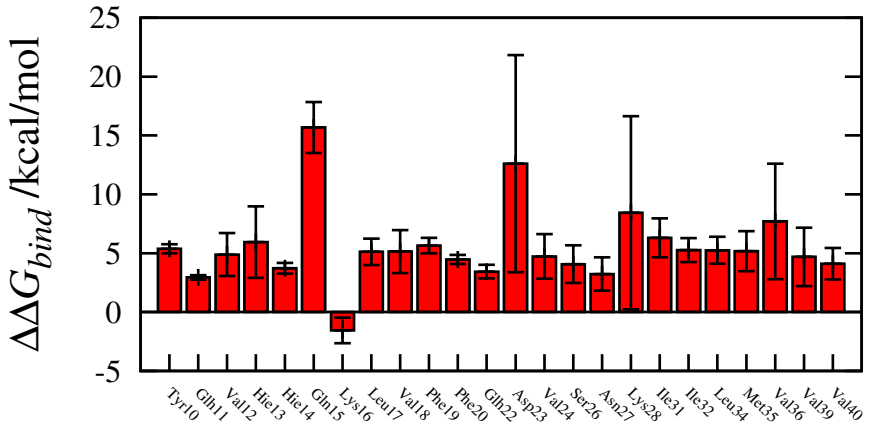
Residue	Method **A**	Method **C**	Method **D**
$Lys28_{C(I)}$	12.5	10.0	13.4
$Lys28_{C(II)}$	9.8	8.8	10.8
$Lys28_{D(I)}$	1.3	0.5	0.5
$Lys28_{D(II)}$	−0.2	−0.6	−0.6

The mutation of residues Glu11, Lys16 and Glu22 to alanine produced a more favourable interfacial binding interaction, compared to that in the wild-type fibril (Figure 3.24).

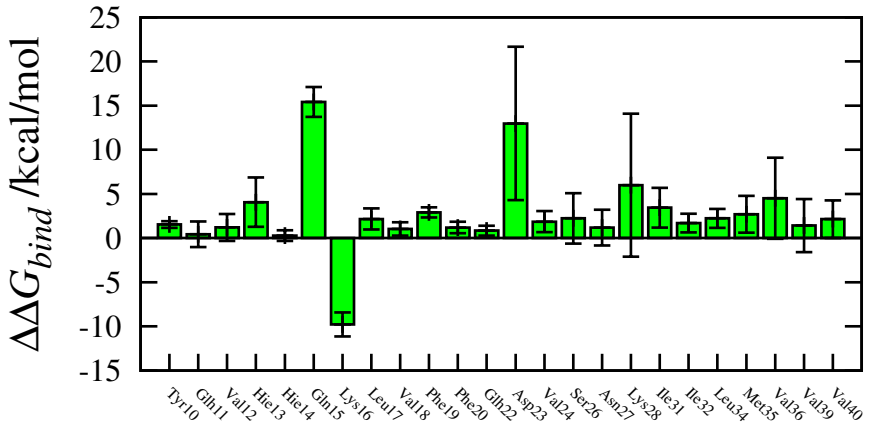
3.3.3 Discussion

3.3.3.1 Interfacial Binding Free Energies

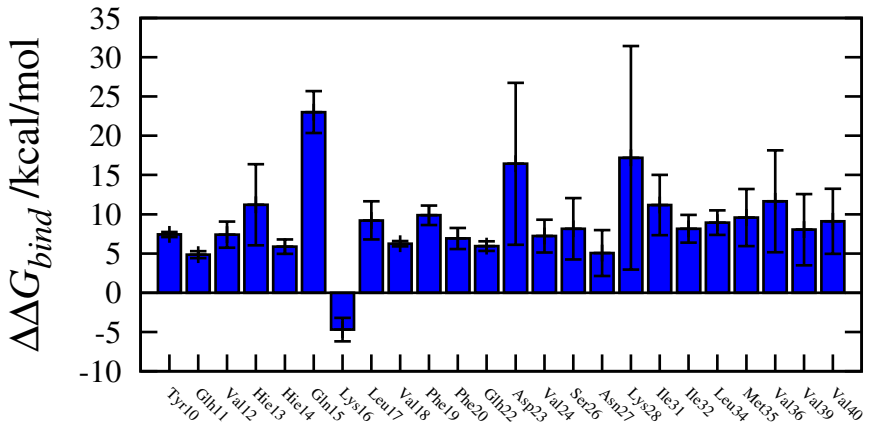
We have applied free energy calculations in order to provide insight into the mechanical stability of the amyloid fibril. Indeed, the 3D structure of the molecular organization of the amyloid fibril has only recently been uncovered through solid state NMR experiments indicating the stacked β-sheet arrangement of Aβ peptides, involving a network of backbone hydrogen bonds (Figure 3.2). As for peptide-fibril interactions, we observe that free energy methods **A** and **D** yield broadly similar trends in binding free energies and associated components; and that method **C** provides rather different predictions, presumably due to errors in its calculation of the repulsive contribution to the nonpolar solvation energy, as discussed in section 3.2.3. Nevertheless, the three methods, applied to the three charge states **1-3**, broadly agree in finding the lateral interface to be more



(a) Method **A**

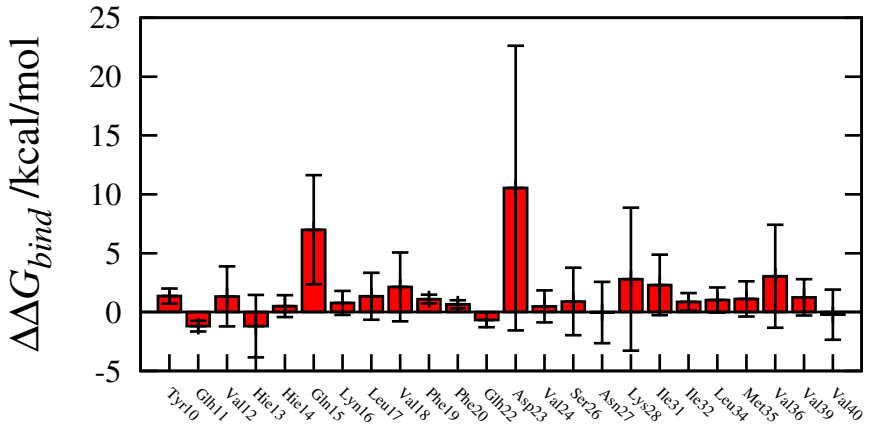


(b) Method **C**

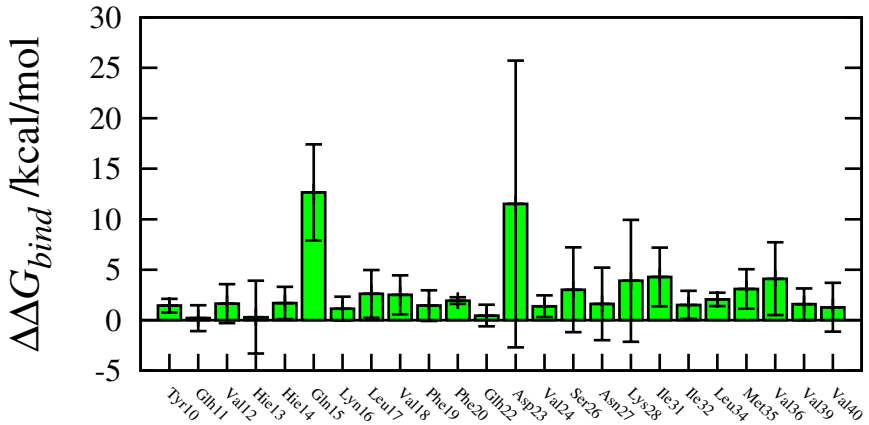


(c) Method **D**

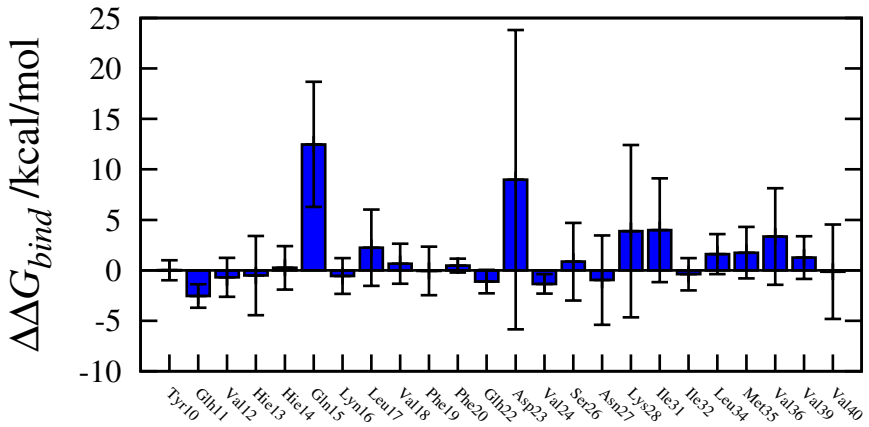
Figure 3.25: Computational alanine scanning of the T_{C-D} interface of the fibril in protonation state **2**. Error bars show standard deviation across four fibril monomers.



(a) Method **A**



(b) Method **C**



(c) Method **D**

Figure 3.26: Computational alanine scanning of the T_{C-D} interface of the fibril in protonation state **2**. Error bars show standard deviation across four fibril monomers.

weakly binding than the transverse interfaces, by an order of 2-3. This suggests that mechanical manipulation of the fibril to dissociate the structure along this lateral interface would be easier than along a transverse interface. Energetic analysis also indicates that the strength of interaction at the transverse interface is not strongly dependent on whether the peptides are situated at the edge or in the middle of the fibril. This agrees with a steered molecular dynamics (SMD) study of the fibril, reported by Raman et al,[96] where individual peptides were computationally dissociated from the fibril, both at the edge and in the middle of the assembly. The forces involved were found to be similar. The study also compared the force required to detach the peptides by pulling in the direction of the transverse interface with that required in the direction along the lateral interface. It was found that dissociation in the latter direction was more favourable than along the transverse interface direction (by a factor of 2). This was due to the large scale and near-simultaneous rupturing of the cooperative backbone hydrogen bonding network by significant peptide pulling in the transverse interface direction; as opposed to the more gradual consequences of breaking fewer hydrogen bonds and some salt bridges on pulling in the orthogonal direction. Although the SMD study did not examine dissociation perpendicular to the direction of our lateral interface, our MM–PBSA calculations suggest that the effects of disrupting the lateral interface would similarly be less severe than breaking interactions across the transverse interface.

3.3.3.2 Computational Alanine Scanning of Fibril $\mathbf{T_{C–D}}$ Interface

In order to understand in more detail the factors governing the stability of the fibril, we have performed computational alanine scanning, for the three protonation states **1** – **3**, using the three nonpolar solvation methods **A**, **C** and **D**.

Within the confines of the free energy models, there appears to be agreement across charge states that residues Gln15 and Asp23 contribute most significantly to fibril stability (Figures 3.24 – 3.26).

We may compare our calculated alanine scanning residue contributions to fibril stability with those experimentally determined in the Wetzel group.[97] Unfortunately it is not possible to compare with experiment for Asp23, as the mutant fibril was not experimentally characterisable. For Gln15, experimental scanning does indicate that Gln15 makes a favourable contribution although by no means the most favourable of either the polar or nonpolar residues in the peptide. Given the large variation in polar residue contributions (and their error bars), we distinguish them from discussion of the contributions of nonpolar residues. The experimentally observed favourable contributions of nonpolar residues 17 – 20 and 31 – 36 reflect their role in stacking to form a β-sheet type structure. Experimentally, of the nonpolar residues, it is found that Phe19, Val24 and Ile31 contribute to stability most prominently, with free energies of 1.5, 1.4 and 2.0 kcal/mol respectively (Figure 3.27). The least favourable contributions of ~0.4 kcal/mol are displayed by residues Val39 and Val40. This trend is reflected across the three fibril charge states and across the three energy methods **A**, **C** and **D**, with Phe19, Val24 and Ile31 among the highest contributing of the nonpolar residues and Val39 and Val40 among the least (Figures 3.18 to 3.20). Interestingly, the range in absolute values of the nonpolar residue contributions between calculation and experiment appear to correspond reasonably well, although this is least true for methods **A** and **D** for protonation state **2** (Figures 3.25a and 3.24c) and method **D** for protonation state **3** (Figure 3.26c).

The contributions of polar residues varied rather more from the findings of Wetzel. This might be anticipated from the greater errors generally associated with im-

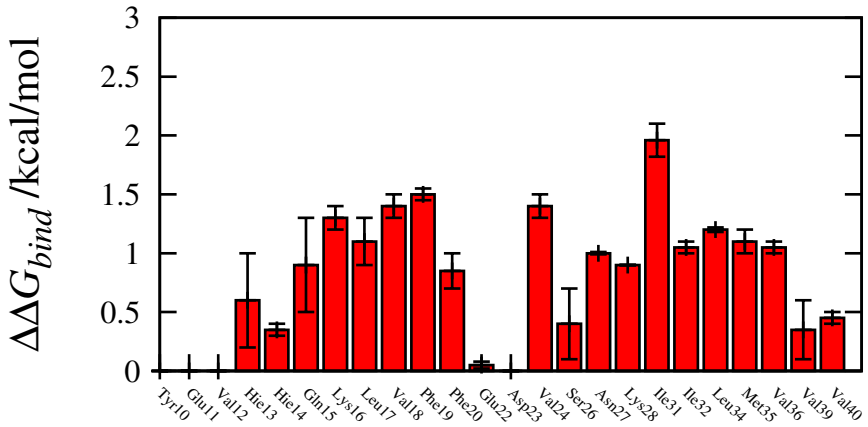


Figure 3.27: Experimental alanine scanning results of Wetzel. Adapted from[97].

plicit solvent esimates of charged solutes. From their alanine scanning, Lys16 in particular is observed to contribute favourably to fibril stability by 1.3 kcal/mol (Figure 3.27). However, in all but the method **A** and **C** estimates for neutral Lys16 (protonation state **3**, Figures 3.26a and 3.26b), this residue is predicted to make an unfavourable contribution to fibril stability. This appears to point once again to sensitivity to orientation of polar residues (particularly when charged) and to issues of adequate sampling of these orientations on the time scale of the simulations. Greater success is found in predicting the favourable contributions of Gln15 and Lys28, although issues in the sampling of the latter residue's interactions have already been highlighted (Table 3.12). We note that the timescale for rotation of surface exposed protein side-chains is typically of the order of 10 – 100 ps. However, for interior side-chains, which may be more representative of some of the interfacial fibril residues examined here, these rotations can occur much less frequently, and can be of the order of 100 μs – 1 s, which is well outside the sampling regime of these 100 ns simulations. However, we also add the caveat that our alanine scanning calculations do not permit relaxation of the alanine mutant fibril in response to the change in side-chain; by contrast, experiments do

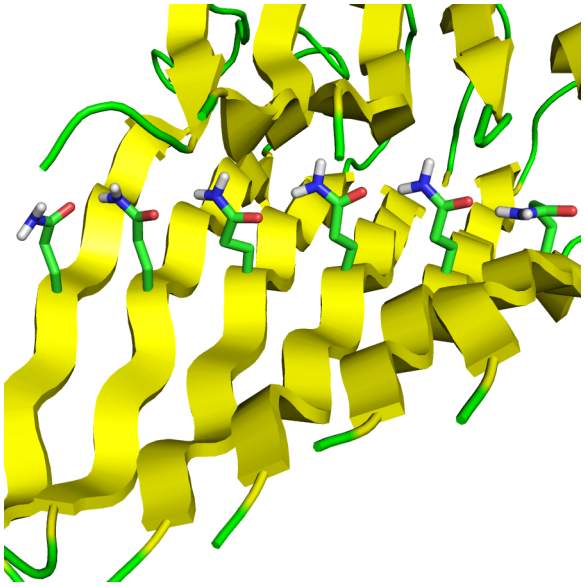


Figure 3.28: Spine of Gln15 residues stabilising the fibril β-sheet face

incorporate the energetic and structural effects of this relaxation, where workable fibril mutants are able to form.

Indeed, Gln15 forms a hydrogen bonded spine that appears to stabilise the β-sheet face of the fibril (Figure 3.28). This is also the region that LPFFD was found to preferably bind (Section 3.2.2).

3.4 Conclusions

From the molecular dynamics simulations of Aβ in complex with the peptides LPFFD and LHFFD, we found that the simulations could indeed distinguish a useful peptide inhibitor from an inactive one that differed at only one residue position. The inactive LHFFD bound less strongly to the fibril than the active peptide, LPFFD, and formed different interactions. LPFFD interacted significantly with the fibril throughout the simulation, while LHFFD, depending on initial location, detached from the fibril, to rebind at a later time point. Cluster analysis of the bound peptide poses suggested that the ability to form inter-monomer hydrogen bonds that could disrupt the β-sheet formation, and especially a key groove-bound orientation of LPFFD, could account for its activity. These results illustrate the promise of MD simulations in compound selection and design procedures, providing an alternate way to estimate the stability of interactions with the target.

We have performed 100 ns explicit solvent MD simulations of amyloid fibril in three putative protonation states. Over the course of these simulations, the least deviation from the initial solid state NMR structures are exhibited by the fibrils in which the Glu11 and Glu22 side-chains are protonated. Subsequent free energy calculations on these trajectories suggest that the weakest fibril interface lies in the lateral rather than transverse direction. We also predict that there is little dependence on whether the lateral interface is at the edge or middle of the fibril, in agreement with reported steered molecular dynamics calculations. Alanine scanning provides residue-level detail into the factors contributing to stability in the fibril. The contributions of nonpolar residues broadly agree with experimental alanine scanning results. However, the important of interactions

between a number of polar residues appears to be overestimated. It is likely that shortcomings in the solvation model and in sampling of side-chains contributions are in part responsible for this discrepancy.

Finally, we note two points concerning the future research of inhibitors of Aβ aggregation. Firstly, we comment that it is tempting to suggest that it may be advantageous to target inhibitors designed to dissociate existing fibril by targeting the weaker lateral interface, although we concede that this is a challenging, shallowly grooved surface in which to obtain such specificity. Secondly, the specifically hydrogen bonded network of Gln15 residues may represent an interesting target site within Aβ fibrils. Although these residues are interior to the fibril, and not exposed to attack by ligands, agents that interact with the solvent exposed groove between His14 and Lys16, as predicted for LPFFD, may be able to disrupt intermonomer interactions in the region.

Chapter 4

Swarm-based Molecular Dynamics

4.1 Swarm Intelligence

Despite the vast conformational space available to them, globular proteins are able to fold rapidly to their unique native geometries.[98] Accurate prediction of these folded states, using only primary sequence information, is an important goal in pharmaceutical science, as knowledge of the structure of protein targets is an important step in rational drug design. Knowledge-based[99] and coarse-graining methods[100] have had some success in predicting protein structure. There are associated drawbacks, however: for example, a dependence on available geometrically similar structures in the former approach; and a lack atomic resolution for the latter, which may lead to omission of important structural features.

Atomistic simulations are ideally placed to provide direct molecular level insights into the structure and dynamics of proteins. However, such approaches are hindered by the simulation timescales required to observe folding events. The free energy surface upon which protein simulations operate is rugged, and characterized by a broad range of barriers. Advanced simulation methods (see Section 2.5) seek to increase the rate at which these barriers are traversed, while maintaining the representative features of the free energy surface; these methods offer the potential to increase the rate at which folding events occur during simulation, allowing the study of larger systems with more complex folding mechanisms.

One route to enhanced sampling that has proved successful in other areas of computational chemistry involves artificial intelligence methods. Genetic algorithms have had a major impact as conformational search tools in protein-ligand docking.[74;75;101] An alternative class of artificial intelligence method is based on swarm intelligence, the emergent behavior observed in nature when social animals, such as swarming insects, flocking birds or schooling fish, act together cooperatively.

In groups, the animals are able to show a greater searching efficiency than they would when acting individually. A concept originally applied computationally in 1989 to cellular robotic systems,[102] the swarm behavior is modeled by a set of simple rules which describe how individual agents act cooperatively within the system.

Recently swarm intelligence approach based on the behavior of ant colonies has proved successful in guiding molecular docking[103] and loop refinement.[104] Another swarm intelligence approach, particle swarm optimization[105] has been used in the development of QSAR models[106] and molecular docking.[107] For the latter, the swarm algorithm exhibited significant improvements in RMSD of pose for 37 protein-ligand complexes, when compared to GOLD (Darwinian genetic algorithm), AutoDock (Larmakian genetic algorithm) and the commonly-used FlexX and DOCK methods.[107]

Huber and van Gunsteren reported a swarm-based molecular dynamics protocol, SWARM-MD,[41] in which a swarm of interacting simulation replicas cooperatively search the conformational space available to the system. In this chapter the implementation and further development of this method will be discussed.

4.2 SWARM-MD

SWARM-MD is a method by which multiple molecular dynamics simulations are made to interact cooperatively. During each step of the dynamics, an additional force is added to each simulation replica, that biases the replica towards the average structure of the whole swarm. The average structure is defined in dihedral

angle space, with the total potential acting on the swarm of M replicas given by:

$$V_{swarm}^{total} = \sum_{j}^{M} A \ \exp\left[-B\left(\frac{1}{N}\sum_{i=1}^{N}\left(\phi_i^j - \langle\phi_i\rangle\right)^2\right)^{\frac{1}{2}}\right] \tag{4.1}$$

where ϕ_i^j is the value of the torsional angle i in replica j, and N is the number of dihedral angles in each replica.

4.2.1 Derivation of SWARM-MD Forces

The total potential acting on the swarm is given by Equation 4.1. Therefore, the potential experienced by swarm member j is:

$$V_{swarm}^{j} = A \ \exp\left[-B\left(\frac{1}{N}\sum_{i=1}^{N}\left(\phi_i^j - \langle\phi_i\rangle\right)^2\right)^{\frac{1}{2}}\right] \tag{4.2}$$

The atomic forces produced by the swarm potential are calculated as the derivative of the potential with respect to the atomic coordinates. These are obtained through chain differentiation. The force acting on atom k in swarm member j is given by:

$$\mathbf{F}_k^j = -\frac{\partial V_{swarm}^j}{\partial \mathbf{r}_k^j} = -\sum_{i}^{N} \frac{\partial V_{swarm}^j}{\partial \phi_i^j}\frac{\partial \phi_i^j}{\partial \mathbf{r}_k^j} \tag{4.3}$$

The derivative of this potential with respect to dihedral angle i in member j is:

$$
\begin{aligned}
\frac{\partial V_j}{\partial \phi_i^j} =& \frac{\partial}{\partial \phi_i^j} A \ \exp\left[-B\left(\frac{1}{N}\sum_{i=1}^{N}\left(\phi_i^j - \langle\phi_i\rangle\right)^2\right)^{\frac{1}{2}}\right] \\
=& A \ \exp\left[-B\left(\frac{1}{N}\sum_{i=1}^{N}\left(\phi_i^j - \langle\phi_i\rangle\right)^2\right)^{\frac{1}{2}}\right] \\
& \times \frac{\partial}{\partial \phi_i^j}\left[-B\left(\frac{1}{N}\sum_{i=1}^{N}\left(\phi_i^j - \langle\phi_i\rangle\right)^2\right)^{\frac{1}{2}}\right] \\
=& A \ \exp\left[-B\left(\frac{1}{N}\sum_{i=1}^{N}\left(\phi_i^j - \langle\phi_i\rangle\right)^2\right)^{\frac{1}{2}}\right] \\
& \times -\frac{B}{2}\left(\frac{1}{N}\sum_{i=1}^{N}\left(\phi_i^j - \langle\phi_i\rangle\right)^2\right)^{-\frac{1}{2}} \\
& \times \frac{\partial}{\partial \phi_i^j}\frac{1}{N}\sum_{i=1}^{N}\left(\phi_i^j - \langle\phi_i\rangle\right)^2 \\
=& A \ \exp\left[-B\left(\frac{1}{N}\sum_{i=1}^{N}\left(\phi_i^j - \langle\phi_i\rangle\right)^2\right)^{\frac{1}{2}}\right] \\
& \times -\frac{B}{2}\left(\frac{1}{N}\sum_{i=1}^{N}\left(\phi_i^j - \langle\phi_i\rangle\right)^2\right)^{-\frac{1}{2}} \\
& \times \frac{2}{N}\left(\phi_i^j - \langle\phi_i\rangle\right) \\
& \times \frac{\partial}{\partial \phi_i^j}\left(\phi_i^j - \langle\phi_i\rangle\right) \\
=& - A \ \exp\left[-B\left(\frac{1}{N}\sum_{i=1}^{N}\left(\phi_i^j - \langle\phi_i\rangle\right)^2\right)^{\frac{1}{2}}\right] \\
& \times \frac{B\left(\phi_i^j - \langle\phi_i\rangle\right)}{N\left(\frac{1}{N}\sum_{i'=1}^{N}\left(\phi_{i'}^j - \langle\phi_i\rangle\right)^2\right)^{\frac{1}{2}}} \\
& \times \left(1 - \frac{\partial}{\partial \phi_i^j}\langle\phi_i\rangle\right)
\end{aligned}
\tag{4.4}
$$

From Equation 4.4 it can be seen that the force on atom k in swarm member

j depends on the derivative of the swarm average of each dihedral angle i with respect to dihedral angle i in swarm member j. At this point, it is necessary to discuss how the swarm average of each dihedral angle is defined.

4.2.1.1 Treatment of Swarm Average

In the original implementation,[41] Huber and van Gunsteren used a simple mean of the unsigned value of each dihedral angle:

$$\langle \phi_i \rangle = \frac{1}{M} \sum_{j=1}^{M} \phi_i^j \tag{4.5}$$

Using this expression, the derivative of the swarm average of dihedral angle i with respect to dihedral angle i in member j is:

$$\begin{aligned} \frac{\partial}{\partial \phi_i^j} \langle \phi_i \rangle &= \frac{\partial}{\partial \phi_i^j} \frac{1}{M} \sum_{j'=1}^{M} \phi_i^{j'} \\ &= \frac{1}{M} \end{aligned} \tag{4.6}$$

Combining Equations 4.4 and 4.6 gives the derivative of the swarm potential acting on swarm member j, with respect to dihedral angle i in swarm member j, as:

$$\frac{\partial V_j}{\partial \phi_i^j} = -A\ \exp\left[-B\left(\frac{1}{N}\sum_{i=1}^{N}\left(\phi_i^j - \langle\phi_i\rangle\right)^2\right)^{\frac{1}{2}}\right] \times \frac{B\left(\phi_i^j - \langle\phi_i\rangle\right)}{N\left(\frac{1}{N}\sum_{i'=1}^{N}\left(\phi_{i'}^j - \langle\phi_i\rangle\right)^2\right)^{\frac{1}{2}}} \times \left(1 - \frac{1}{M}\right) \tag{4.7}$$

In this work we propose using an alternative average definition, which can account for the sign of the dihedral angles, and their periodic nature. By this definition, the average value of a set of dihedral angles is given by:[108]

$$\langle\phi_i\rangle = \arctan\left(\frac{\sum_{j=1}^{M}\sin\phi_i^j}{\sum_{j'=1}^{M}\cos\phi_i^{j'}}\right) \tag{4.8}$$

where arctan is the two argument arctangent function, which returns values in the range $-180° < x < 180°$. The derivative of this expression is:

$$\frac{\partial}{\partial\phi_i^j}\langle\phi_i\rangle = \frac{\partial}{\partial\phi_i^j}\arctan\left(\frac{\sum_{j'=1}^{M}\sin\phi_i^{j'}}{\sum_{j''=1}^{M}\cos\phi_i^{j''}}\right)$$

$$
\begin{aligned}
=&\frac{1}{1+\left(\frac{\sum_{j'=1}^{M}\sin\phi_i^{j'}}{\sum_{j''=1}^{M}\cos\phi_i^{j''}}\right)^2} \\
&\times\left(\frac{\left(\cos\phi_i^j\sum_{j'''=1}^{M}\cos\phi_i^{j'''}\right)+\left(\sin\phi_i^j\sum_{j''''=1}^{M}\sin\phi_i^{j''''}\right)}{\left(\sum_{j'''''=1}^{M}\cos\phi_i^{j'''''}\right)^2}\right) \\
=&\frac{\left(\cos\phi_i^j\sum_{j'=1}^{M}\cos\phi_i^{j'}\right)+\left(\sin\phi_i^j\sum_{j''=1}^{M}\sin\phi_i^{j''}\right)}{\left(\sum_{j'''=1}^{M}\cos\phi_i^{j'''}\right)^2+\left(\sum_{j''''=1}^{M}\sin\phi_i^{j''''}\right)^2}
\end{aligned}
\tag{4.9}
$$

Combining Equations 4.4 and 4.9 we get:

$$
\begin{aligned}
\frac{\partial V_j}{\partial\phi_i^j}=&-A\ \exp\left[-B\left(\frac{1}{N}\sum_{i=1}^{N}\left(\phi_i^j-\langle\phi_i\rangle\right)^2\right)^{\frac{1}{2}}\right] \\
&\times\frac{B\left(\phi_i^j-\langle\phi_i\rangle\right)}{N\left(\frac{1}{N}\sum_{i'=1}^{N}\left(\phi_{i'}^j-\langle\phi_i\rangle\right)^2\right)^{\frac{1}{2}}} \\
&\times\left(1-\frac{\left(\cos\phi_i^j\sum_{j'=1}^{M}\cos\phi_i^{j'}\right)+\left(\sin\phi_i^j\sum_{j''=1}^{M}\sin\phi_i^{j''}\right)}{\left(\sum_{j'''=1}^{M}\cos\phi_i^{j'''}\right)^2+\left(\sum_{j''''=1}^{M}\sin\phi_i^{j''''}\right)^2}\right)
\end{aligned}
\tag{4.10}
$$

In discussions that follow, we refer to the average given by Equation 4.5 as the non-periodic average definition, and that given by Equation 4.8 as the periodic average definition.

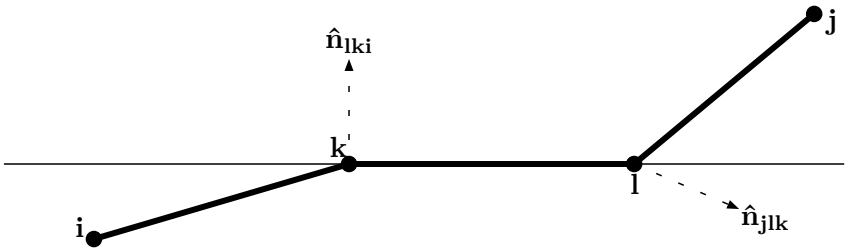


Figure 4.1: Normal vectors used to define the dihedral between four atoms.

4.2.1.2 Derivative of Dihedral Angle with Respect to Atomic Position

As shown in Equation 4.3, in order to calculate the atomic forces, we need to know how a dihedral angle changes with respect to changes in the position of the four atoms which define the angle. Figure 4.1 shows how the dihedral angle ϕ can be defined by the unit vectors $\hat{\mathbf{n}}_{\mathbf{lki}}$, normal to the plane of ijk and $\hat{\mathbf{n}}_{\mathbf{jlk}}$, normal to the plane of jlk. The normal vector $\hat{\mathbf{n}}_{\mathbf{lki}}$ can be obtained from the vector product of $\mathbf{r}_{\mathbf{ki}}$ and $\mathbf{r}_{\mathbf{kl}}$, and $\hat{\mathbf{n}}_{\mathbf{jlk}}$ from the vector product of $\mathbf{r}_{\mathbf{lk}}$ and $\mathbf{r}_{\mathbf{jk}}$:

$$\hat{\mathbf{n}}_{\mathbf{lki}} = \frac{\mathbf{r}_{\mathbf{ki}} \times \mathbf{r}_{\mathbf{kl}}}{|\mathbf{r}_{\mathbf{ki}} \times \mathbf{r}_{\mathbf{kl}}|} \tag{4.11}$$

$$\hat{\mathbf{n}}_{\mathbf{jlk}} = \frac{\mathbf{r}_{\mathbf{lk}} \times \mathbf{r}_{\mathbf{kj}}}{|\mathbf{r}_{\mathbf{lk}} \times \mathbf{r}_{\mathbf{kj}}|} \tag{4.12}$$

The cosine of dihedral angle ϕ can then be obtained from the scalar product:

$$\cos\phi = (\hat{\mathbf{n}}_{\mathbf{lki}} \cdot \hat{\mathbf{n}}_{\mathbf{jlk}}) \tag{4.13}$$

The derivative of ϕ, with respect to the position of atom i can be obtained by chain differentiation:

$$\frac{\mathrm{d}\phi}{\mathrm{d}\mathbf{r}_{\mathbf{i}}} = \frac{\mathrm{d}\phi}{\mathrm{d}\cos\phi}\frac{\mathrm{d}\cos\phi}{\mathrm{d}\mathbf{r}_{\mathbf{i}}} = \frac{-1}{\cos\phi}\frac{\mathrm{d}\cos\phi}{\mathrm{d}\mathbf{r}_{\mathbf{i}}} \tag{4.14}$$

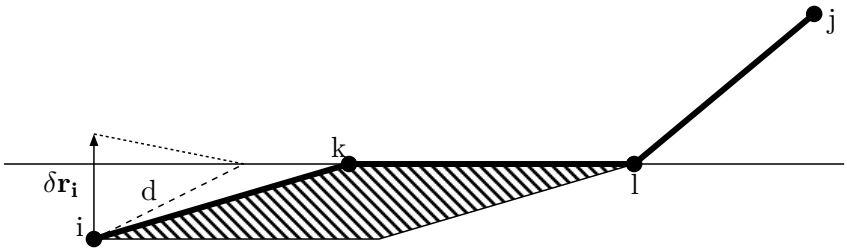


Figure 4.2: Dihedral angle force calculation as described by Bekker et al

The presence of the $-1/\cos\phi$ term causes a discontinuity when $\phi = 180$ or 0. An alternative definition of this derivative, which avoids the discontinuity was shown by Bekker et al.[109] Their method, based on geometric arguments concerning the forces acting on the atoms, is shown in Figure 4.2.

The force acting on atom i, acts against the potential which causes it, so the direction of the force must be perpendicular to the plane of lki because any movement within this plane results in no change in the potential. The force must therefore be acting in the direction of $\hat{\mathbf{n}}_{\mathbf{lki}}$, and so

$$\mathbf{F_i} = |\mathbf{F_i}|\hat{\mathbf{n}}_{\mathbf{lki}} \tag{4.15}$$

The shaded area in Figure 4.2 is a parallelogram of base $\mathbf{r_{kl}}$ and height d, where d is the distance from atom i to the extension of vector $\mathbf{r_{kl}}$, along a direction perpendicular to $\mathbf{r_{kl}}$. Its area can be found from the magnitude of the vector product of the vectors $\mathbf{r_{ki}}$ and $\mathbf{r_{kl}}$, Therefore we can write:

$$d = \frac{|\mathbf{r_{ki}} \times \mathbf{r_{kl}}|}{r_{kl}} \tag{4.16}$$

For a small displacement of atom i, $\delta\mathbf{r_i}$, producing a small change in ϕ of $\delta\phi$:

$$\tan(\delta\phi) = \frac{\delta\mathbf{r_i}}{d} = \delta\mathbf{r_i}\frac{r_{kl}}{|\mathbf{r_{ki}} \times \mathbf{r_{kl}}|} \tag{4.17}$$

so for an infinitesimal $\delta\phi$

$$\frac{\delta\phi}{\delta\mathbf{r_i}} = \frac{r_{kl}}{|\mathbf{r_{ki}} \times \mathbf{r_{kl}}|} \tag{4.18}$$

If we define the vector product of $\mathbf{r_{ki}}$ and $\mathbf{r_{kl}}$ as $\mathbf{m}$, we can write the magnitude of the force on i as:

$$|\mathbf{F_i}| = \left|-\frac{\mathrm{d}V(\phi)}{\mathrm{d}\phi}\right| \frac{\delta\phi}{\delta\mathbf{r_i}} = \left|-\frac{\mathrm{d}V(\phi)}{\mathrm{d}\phi}\right| \frac{r_{kl}}{|\mathbf{m}|} \tag{4.19}$$

Combining this with Equation 4.15, we can write:

$$\mathbf{F_i} = -\frac{\mathrm{d}V(\phi)}{\mathrm{d}\phi}\frac{\delta\phi}{\delta\mathbf{r_i}} = -\frac{\mathrm{d}V(\phi)}{\mathrm{d}\phi}\frac{\mathbf{r_{kl}}}{|\mathbf{m}|}\hat{\mathbf{m}} = -\frac{\mathrm{d}V(\phi)}{\mathrm{d}\phi}r_{kl}\frac{\mathbf{m}}{|\mathbf{m}|^2} \tag{4.20}$$

Defining the vector product of $\mathbf{r_{lk}}$ and $\mathbf{r_{jk}}$ as $\mathbf{n}$, it can similarly be shown that the force acting on atom j is:

$$\mathbf{F_j} = -\frac{\mathrm{d}V(\phi)}{\mathrm{d}\phi}\frac{\delta\phi}{\delta\mathbf{r_i}} = -\frac{\mathrm{d}V(\phi)}{\mathrm{d}\phi}r_{kl}\frac{\mathbf{n}}{|\mathbf{n}|^2} \tag{4.21}$$

For the forces on the central two atoms k and l, Bekker et al[109] showed that the forces could be described by:

$$\mathbf{F_k} = -\mathbf{F_i} + \frac{\mathbf{r_{ik}} \cdot \mathbf{r_{lk}}}{r_{lk}^2}\mathbf{F_i} - \frac{\mathbf{r_{lj}} \cdot \mathbf{r_{lk}}}{r_{lk}^2}\mathbf{F_j} \tag{4.22}$$

$$\mathbf{F_l} = -\mathbf{F_j} - \frac{\mathbf{r_{ik}} \cdot \mathbf{r_{lk}}}{r_{lk}^2}\mathbf{F_i} - \frac{\mathbf{r_{lj}} \cdot \mathbf{r_{lk}}}{r_{lk}^2}\mathbf{F_j} \tag{4.23}$$

4.2.2 Implementation into Amber 9

The swarm potential, as described above, was implemented into the *sander* module of Amber 9.[76] The flow of information within *sander* is summarised in Figure 4.3. When the program is called, control passes to the *multisander* routine in the *sander* source code. *Multisander* is able to initiate multiple instances of the *sander* subroutine. This allows multiple replica simulations to be performed simultaneously, as would be required for a replica exchange simulation.[31] Each instance of *sander* then reads in the input coordinates, parameters and control information from input files. The control information describes how the simulation is to be run. *Sander* then passes control to the main molecular dynamics routine *runmd*, which performs the simulation.

Firstly, *runmd* sets the initial velocities of the atoms in the simulation, if these are not read in with the coordinate information, then performs a loop over the time steps of the simulation. At each time step, *runmd* calls the subroutine *force*, which calculates the atomic forces and potential energy produced by the molecular mechanics force field potentials (see Section 2.1.2). Control then passes back to *runmd*, where the forces are used to integrate the simulation trajectory. A number of other routines are also called by *runmd* to perform additional operations, such as temperature or pressure regulation[26] or SHAKE[79] constraint corrections, as required by the control information. Coordinate, velocity and energy information will also be printed to output files, if required on the current time step. Once the loop over time steps is complete, final results data is written to the output files.

In our modified version of *sander*, an additional subroutine is called by *force* to calculate the forces due to the swarm potential, as discussed in Section 4.2.1. In this subroutine, each simulation replica calculates the the value of each of

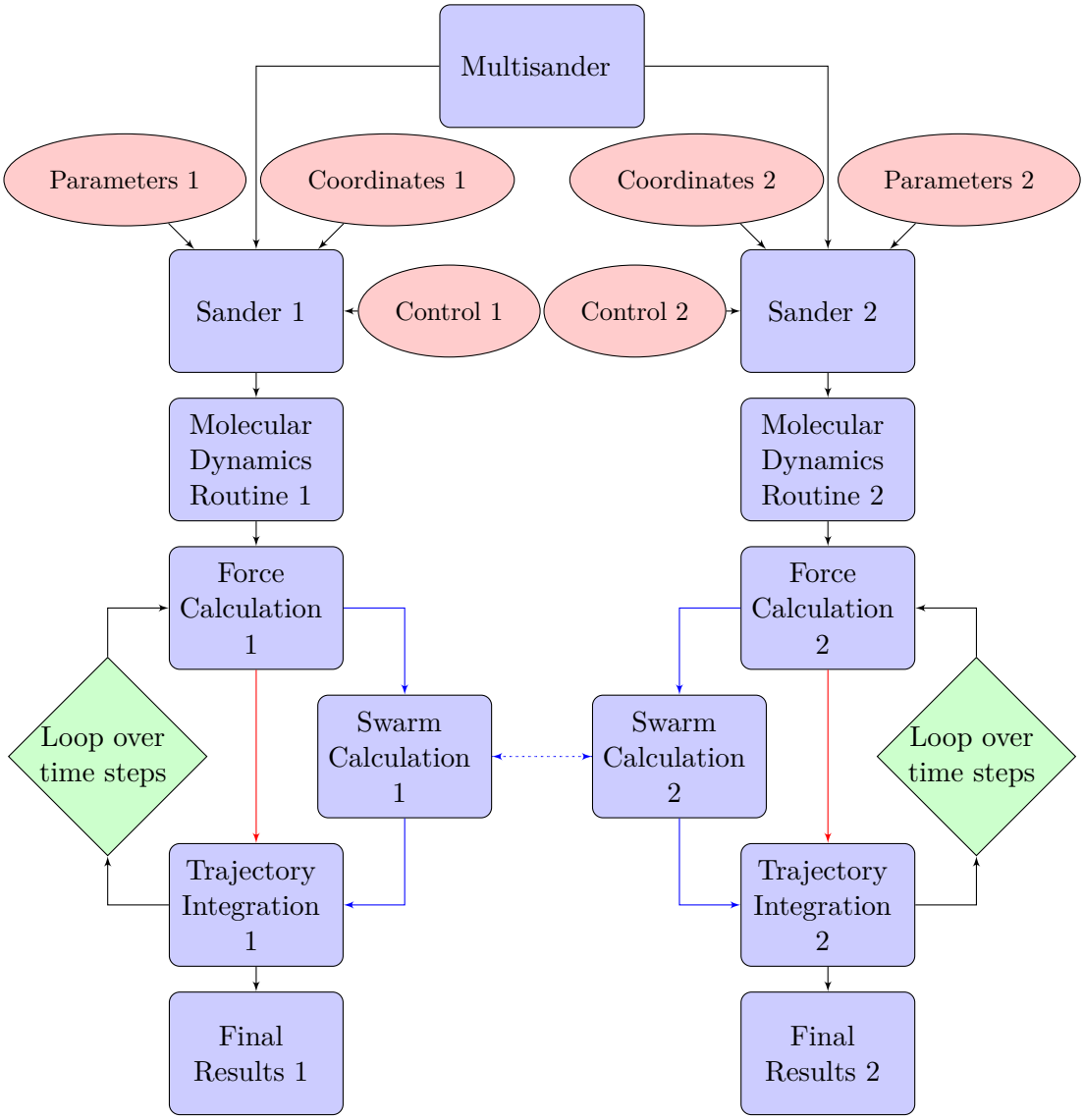


Figure 4.3: Flow diagram showing the flow of information in *sander* during a two-replica molecular dynamics simulation. The red arrow represents the unmodified implementation; while the blue arrows shows the modified SWARM-MD implementation. The dotted arrow between the two swarm calculations depicts the passing of information between the two simulation replicas.

its dihedral angles, then the replicas communicate these values with each other in order to calculate the average values. The resulting swarm forces are then added to the total atomic forces, due to the force field terms, and these are used to integrate the simulation trajectory. The swarm forces can be calculated using either the non-periodic dihedral angle average definition of Huber and van Gunsteren,[41] or the periodic average definition proposed in Section 4.2.1.1.

4.2.2.1 Validation of Swarm Implementation by Finite Difference Calculations

The algorithm was validated by performing a series of finite difference calculations. For model systems, the x, y and z coordinates of each atom in the system were altered by $\pm\Delta r$ where $\Delta r = 0.00005$. The finite difference force F acting on atom i, in coordinate r, was calculated according to:

$$F_r^i = -\left[\frac{V_{swarm}\left(r+\Delta r\right) - V_{swarm}\left(r-\Delta r\right)}{2\Delta r}\right] \tag{4.24}$$

The forces obtained by finite difference were compared to the analytical forces calculated within Amber. Excellent agreement was observed, for example the Figure 4.4 shows the agreement for the capped tyrosine residue Ac-Tyr-NMe ($r^2 = 1.00$).

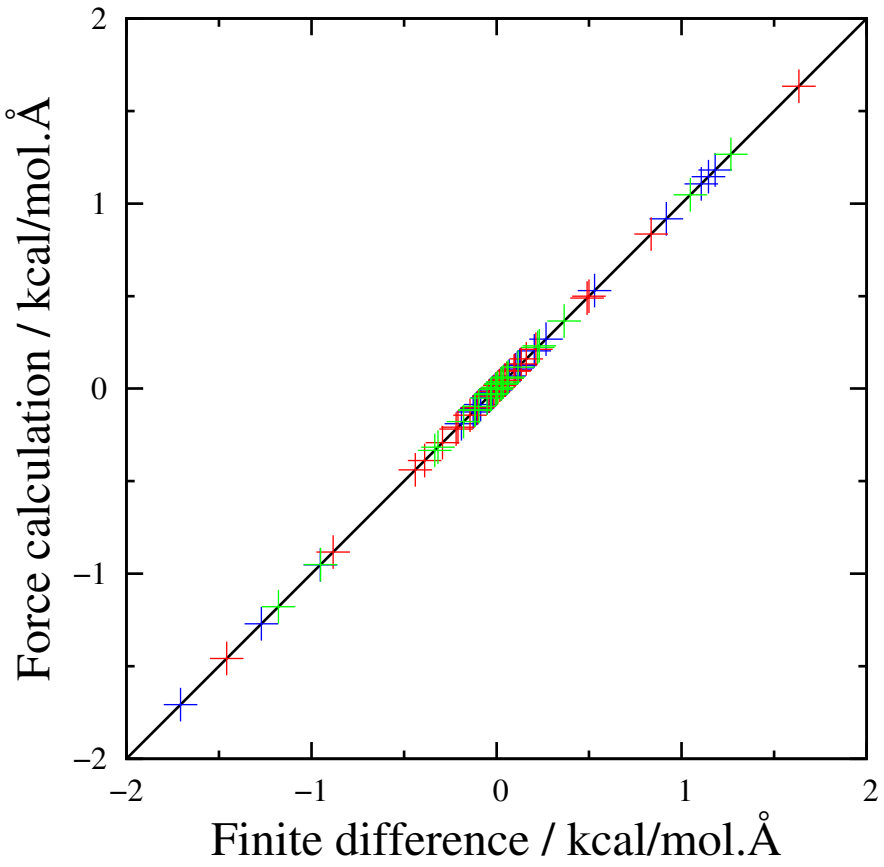


Figure 4.4: Comparison of forces calculated by SWARM-MD in Amber, and forces calculated from finite difference of swarm energy (red, green and blue show the x, y and z components of the forces respectively).

4.3 Predicting Native States Through Swarm-Based Simulated Annealing

NOTE: *The work outlined in this section has been published previously in the Journal of Chemical Theory and Computation.*[2]

4.3.1 Test Systems

In this section we examine the ability of SWARM-MD to identify the native conformations of polyalanine (Ac-$(Ala)_{11}$-NH_2) in the gas phase, a peptide called AEK17 in aqueous solution (Ac-Ala-$(Glu\text{-}Ala\text{-}Ala\text{-}Ala\text{-}Lys)_3$-Ala-$NH_2$) and the 20-residue Trp-cage mini-protein (sequence: N_{20}LYIQWLKDGGPSSGRPPPS$_{39}$), again in aqueous solution.[110] Polyalanine and AEK17 both form α-helical structures;[111;112] while Trp-cage consists of an α-helix (residues 20–28), a short 3_{10} helix (residues 30–33), and a polyproline II helix (residues 36–38).

4.3.2 Methods

All simulations were performed using the modified version of *sander* from Amber 9[76] described in section 4.2.2. The Amber *ff99SB* force field of Simmerling and coworkers was employed and a simulation time step of 1 fs was used. Temperature was controlled using Langevin dynamics.[27] The calculation of non-bonded interactions was not truncated at any distance, and all bonds involving hydrogen atoms were constrained using the SHAKE algorithm.[79] Initial, fully extended ($\phi = \psi = 180°$), conformations of each test peptide were generated using the

tleap module in Amber 9. In their initial paper,[41] Huber and van Gunsteren used a value of 0.8 rad^{-1} for parameter B in Equation 4.1. This value was kept fixed in all simulations, however the effect that the strength of the swarm potential has on the performance of the simulated annealing simulations was monitored by repeating the simulations with differing values of parameter A, as described in sections 4.3.2.1 – 4.3.2.3. Post-simulation structural analyses of the simulation trajectories were performed using *ptraj* from Amber 10[80]. Secondary structure is assigned by *ptraj* using the DSSP algorithm.[113]

4.3.2.1 Polyalanine

The *tleap* generated initial structure of polyalanine was geometry optimised in the gas phase (ie without any treatment of solvent shielding) using *sander*, to remove any initial steric clashes that may destabilise the molecular dynamics simulations. The structure was subjected to 500 steps of steepest descent minimisation, followed by 1000 steps of conjugate gradient minimisation. 20 temperature equilibrated polyalanine structures were generated by slowly heating the minimised geometry from 30 K to 300 K, in the gas phase, over 20 ps, with no swarm forces present in the system. The collision frequency used for Langevin dynamics[27] temperature regulation was 5.0 ps^{-1}. Different random number generator seeds were used for each of the 20 heating simulations, giving different Boltzmann-weighted initial atomic velocities.

These 20 equilibrated polyalanine structures and velocities were used as the initial conditions for each 1.2 ns 20 replica simulated annealing simulation. Coordinate snapshots were recorded at 400 fs intervals, during each simulation trajectory. Simulated annealing was performed in in the absence of any swarm forces (ie $A =$

0.0 kcal/mol); with a non-periodic swarm average definition (as defined in section 4.2.1.1) and swarm potential strengths $A = -0.25$, -0.5, -1.0, -2.0, -5.0, -10.0, -15.0, -20.0, -25.0, -50.0 and -100.0 kcal/mol; and with a periodic swarm average definition and swarm potentials of strength $A = -25.0$, -50.0 and -100.0 kcal/mol.

The same annealing temperature profile was used for each simulation (Table 4.1). During each temperature phase, the simulation target temperature at each time step was linearly interpolated from the initial and final temperatures of that phase and the Langevin dynamics collision frequency of 5.0 ps^{-1} was used for temperature regulation.

Table 4.1: Simulated annealing heating schedule for polyalanine, showing initial and final temperatures (K) for each heating stage (ns)

Time	Initial Temp	Final Temp
$0.0 - 0.1$	300	1000
$0.1 - 0.2$	1000	1000
$0.2 - 1.2$	1000	10

The trajectories obtained through simulation were analysed using *ptraj*[80] to calculate the fractional helicity[113] of each snapshot.

4.3.2.2 AEK17

As in the case of polyalanine, the initial extended structure of AEK17 was optimised through 500 steps of steepest descent and 1000 steps of conjugate gradient minimisation. This time, aqueous solvent was modelled using the generalised Born method of Hawkins, Cramer and Truhlar[14;17;18] implemented in Amber 9. The optimised structure was used as the starting point for 20 heating simulations, in which the temperature was raised from 30 to 300 K over 20 ps. A

collision frequency of 5.0 ps^{-1} was used for Langevin dynamics.[27] The final coordinates and velocities from these simulations were used as the starting points for a series of 12 ns simulated annealing simulations, using the annealing schedule shown in Table 4.2, with the temperatures regulated via Langevin dynamics[27] with a collision frequency of 5.0 ps^{-1}. Annealing was performed in the absence of any swarm potential, and with periodic and non-periodic swarm potentials of strengths $A = -25.0$, -50.0 and -100.0 kcal/mol.

Table 4.2: Simulated annealing heating schedule for AEK17, showing initial and final temperatures (K) for each heating stage (ns)

Time	Initial Temp	Final Temp
$0.0 - 0.3$	300	650
$0.3 - 0.5$	650	650
$0.5 - 2.5$	650	450
$2.5 - 4.5$	450	350
$4.5 - 9.0$	350	250
$9.0 - 12.0$	250	250

Post-simulation analyses were performed using *ptraj*,[80] to determine the fractional helicity[113] and number of native and non-native salt bridges present in each snapshot.

4.3.2.3 Trp-cage

The initial extended structure of Trp-cage was optimised through 500 steps of steepest descent and 1000 steps of conjugate gradient minimisation, with the generalised Born method of Hawkins, Cramer and Truhlar[14;17;18] used to describe the aqueous solvent. 20 heating simulations were performed, beginning from the optimised structure, with the simulation temperature raised from 30 to 325 K, using a Langevin dynamics collision frequency of 2.0 ps^{-1}. Two simulated

annealing schedules were used, a 25 ns protocol shown in Table 4.3, and a longer 40 ns protocol shown in Table 4.4. Simulation temperature was controlled using a Langevin dynamics[27] collision frequency of 1.0 ps^{-1} in during all simulations. The annealing was performed without any swarm potential, and with a periodic swarm potential of strength -50.0 kcal/mol. Following each simulation, post-simulation analyses were performed using *ptraj*.[80]

Table 4.3: Simulated annealing heating schedule A for Trp-cage, showing initial and final temperatures (K) for each heating stage (ns)

Time	Initial Temp	Final Temp
$0.0 - 0.3$	325	450
$0.3 - 1.0$	450	450
$1.0 - 25.0$	450	250

Table 4.4: Simulated annealing heating schedule B for Trp-cage, showing initial and final temperatures (K) for each heating stage (ns)

Time	Initial Temp	Final Temp
$0.0 - 0.3$	325	450
$0.3 - 1.0$	450	450
$1.0 - 10.0$	450	390
$10.0 - 40.0$	390	300

Once the annealing was complete, the swarm potential was tapered to zero, by slowly changing the value of parameter A from -50.0 to 0.0 kcal/mol over 2 ns, then performing a 3 ns simulation with the temperature fixed at 300 K, and no swarm potential affecting the dynamics. To keep the total simulation time the same in the independent simulations, a period of 5 ns constant temperature simulation was performed.

4.3.2.4 Computational Resources

Each of the 20 member simulations of polyalanine took approximately 3 hours to run on 20 CPU cores. A total of 15 of these simulations were performed, giving a total walltime of 45 hours on 20 cores. The 20 swarm member AEK17 simulations required around 50 hours each on 20 cores, and seven such simulations were performed. Two 25 ns simulations were performed with 20 Trp-cage replicas, these each took about 75 hours to complete on 20 processor cores; while the two 40 ns simulations took around 130 hours on 20 cores. The final 5 ns relaxation simulations took about 15 hours on 20 cores.

4.3.3 Results

4.3.3.1 Polyalanine

With no swarm potential, the 20 independent simulation replicas of polyalanine folded to an average fractional helicity of 29% (Figure 4.5). Five of the replicas folded to fully α-helical structures. Note that a replica is defined as being fully folded if all non-terminal residues are assigned an α-helical conformations in the final annealed structure, this corresponds to a helical content of greater than 82 %. Using non-periodic swarm potentials (see Section 4.2.1.1) of strengths $A = -25.0$, -50.0 and -100.0 kcal/mol, the performance of swarm-based simulations decreased. The final folded fractional helicities falling to 14%, 23% and 9% respectively, and the number of folded replicas falling to one, two and one respectively. Using weaker swarm potentials of strengths $A = -5.0$, -10.0, -15.0 and -20.0 kcal/mol (Figure 4.6), the presence of the swarm potential had either a negative effect on folding performance, or no noticeable effect, compared to

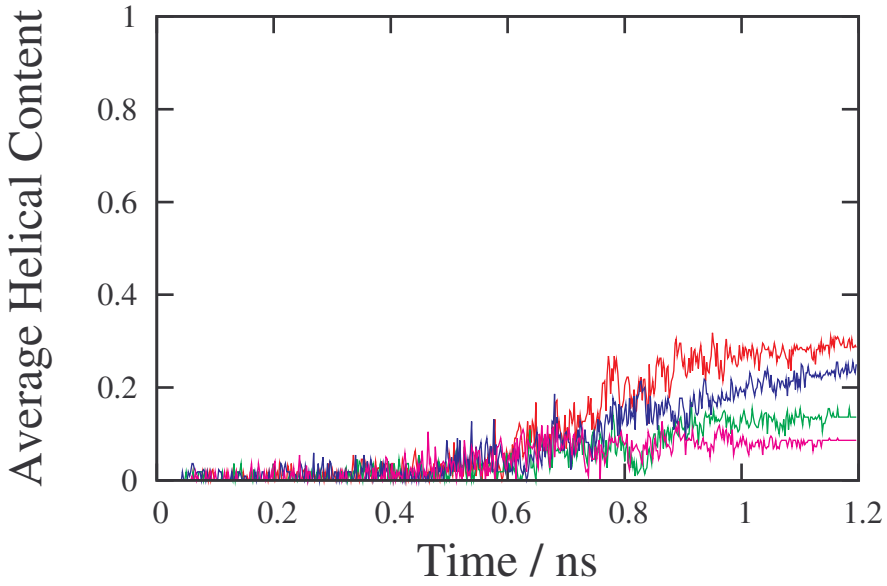


Figure 4.5: Average fractional helicity across 20 simulation replicas during simulated annealing of polyalanine, with no swarm potential present (red) and non-periodic swarm potentials of strength $A = -25.0$ kcal/mol (green), $A = -50.0$ kcal/mol (blue) and $A = -100.0$ kcal/mol (pink).

the independent simulations. In these cases, the final average fractional helicities were 28%, 29%, 16% and 16%, and the number of folded replicas were five, five, two and one respectively.

Using periodic swarm potentials of strengths $A = -25.0$, -50.0 and -100.0 kcal/mol, the final fractional helicity increased in all cases (Figure 4.7), rising to 39%, 75% and 91% respectively. However, the number of fully folded replicas in the $A = -25.0$ kcal/mol simulation did fall to four. The number of fully folded replicas increased to 15 in the -50.0 kcal/mol simulation, while all replicas folded to fully helical conformations during the -100.0 kcal/mol simulation.

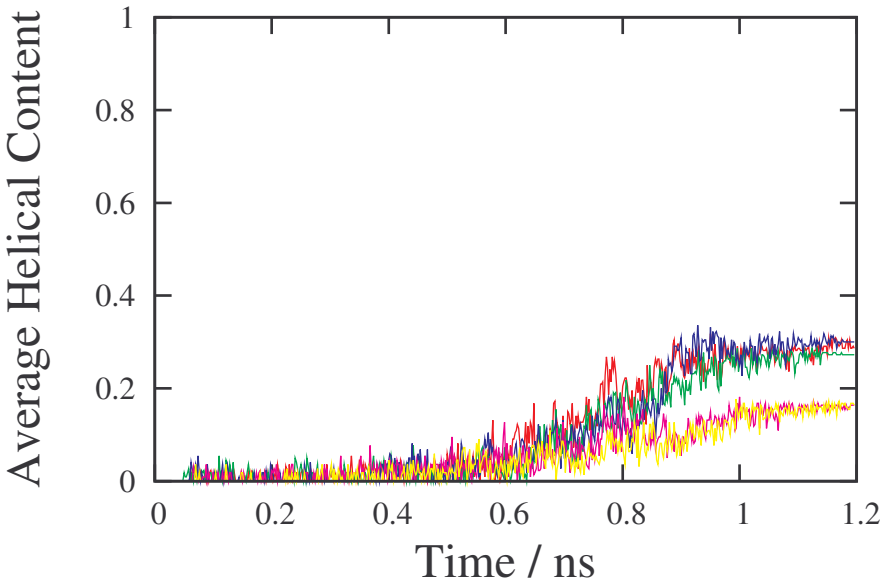


Figure 4.6: Average fractional helicity across 20 simulation replicas during simulated annealing of polyalanine, with no swarm potential present (red) and non-periodic swarm potentials of strength $A = -5.0$ kcal/mol (green), $A = -10.0$ kcal/mol (blue), $A = -15.0$ kcal/mol (pink) and $A = -20.0$ kcal/mol (yellow).

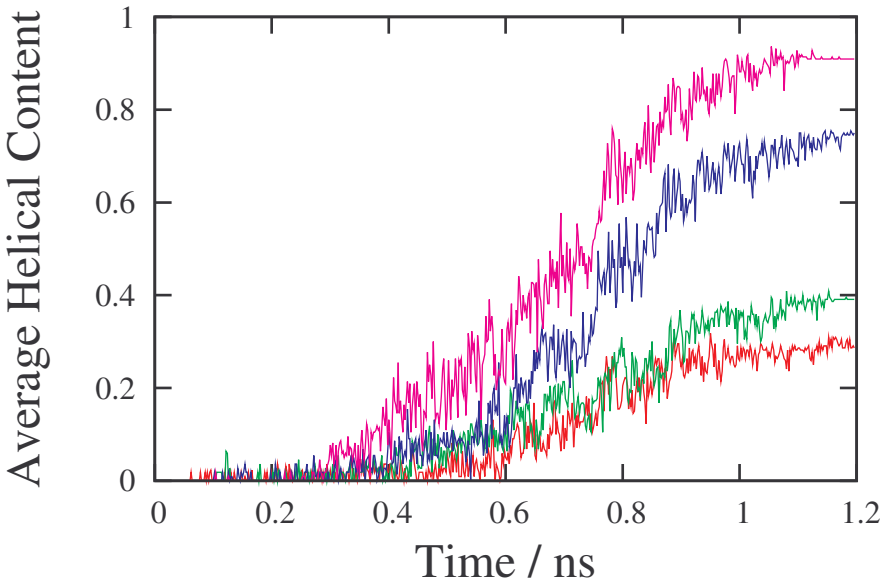


Figure 4.7: Average fractional helicity across 20 simulation replicas during simulated annealing of polyalanine, with no swarm potential present (red) and periodic swarm potentials of strength $A = -25.0$ kcal/mol (green), $A = -50.0$ kcal/mol (blue) and $A = -100.0$ kcal/mol (pink).

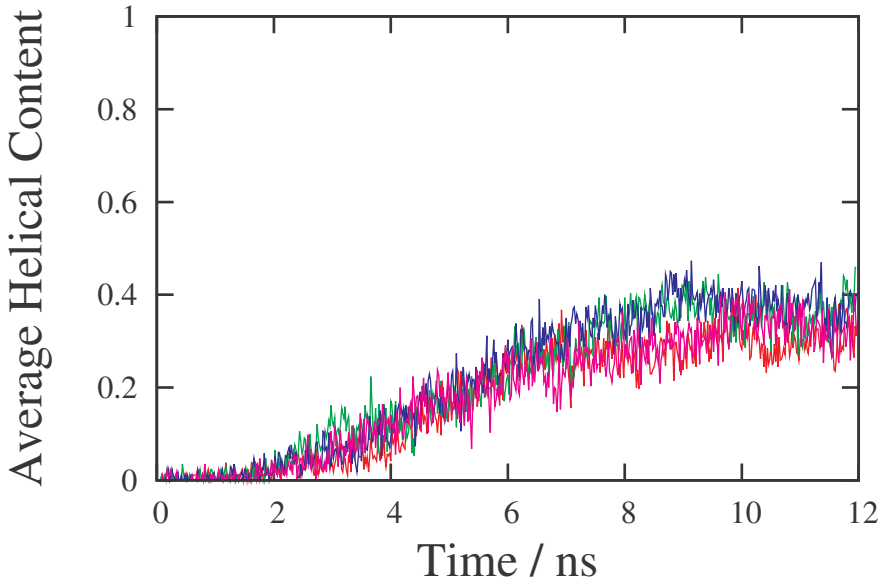


Figure 4.8: Average fractional helicity across 20 simulation replicas during simulated annealing of AEK17, with no swarm potential present (red) and with non-periodic swarm potentials of strength $A = -25.0$ kcal/mol (green), $A = -50.0$ kcal/mol (blue) and $A = -100.0$ kcal/mol (pink).

4.3.3.2 AEK17

Following the simulated annealing protocol outlined in section 4.3.2.2, the 20 independent replicas of AEK17 folded to an average fractional helicity of 30%, averaged over the final 2 ns of constant temperature dynamics (Figure 4.8). Using a non-periodic swarm definition and swarm potentials of strengths $A = -25.0$, $A = -50.0$ and $A = -100.0$ kcal/mol, AEK17 folded to structures with average helicities of 36%, 38% and 33% respectively.

During the final 2 ns of simulation, the average number of native salt bridges (see section 4.3.1) observed across 20 independent replicas was 1.5 (Figure 4.9), while the average number of non-native salt bridges observed was 1.1 (Figure 4.10). The presence of non-periodic swarm potentials in the dynamics of the simulations had only a small effect on the observed salt bridge occupancies. With

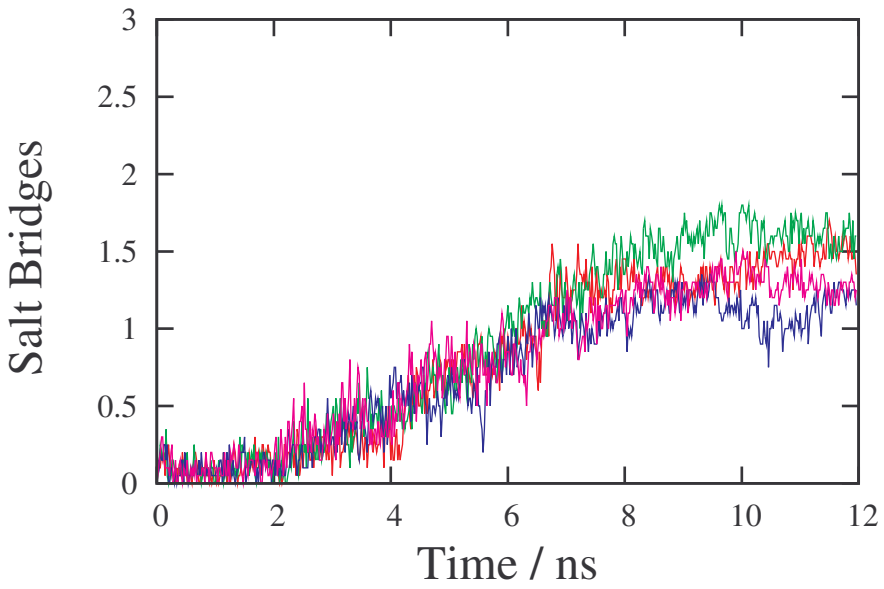


Figure 4.9: Average number of native salt bridges across 20 simulation replicas during simulated annealing of AEK17, with no swarm potential present (red) and with non-periodic swarm potentials of strength $A = -25.0$ kcal/mol (green), $A = -50.0$ kcal/mol (blue) and $A = -100.0$ kcal/mol (pink).

swarm potentials of strengths $A = -25.0$, $A = -50.0$ and $A = -100.0$ kcal/mol, the average numbers of native salt bridges observed were 1.6, 1.1 and 1.3; while the average numbers of non-native salt bridges observed were 1.1, 1.0 and 1.4 respectively.

More specifically, defining a simulation replica as correctly folded if it displays an average fractional helicity of greater than 70% and an average native salt bridge occupancy of greater than 2.0, over the final 2 ns of simulation, two of the 20 independent replicas achieved a correctly folded conformation. With a swarm potential of strength $A = -25.0$ kcal/mol, three simulation replicas folded; with a potential of $A = -50.0$ kcal/mol, no simulation replicas met the criteria above; while with $A = -100.0$ kcal/mol, two replicas folded (Table 4.5).

The use of periodic swarm potentials of strengths $A = -25.0$ kcal/mol and

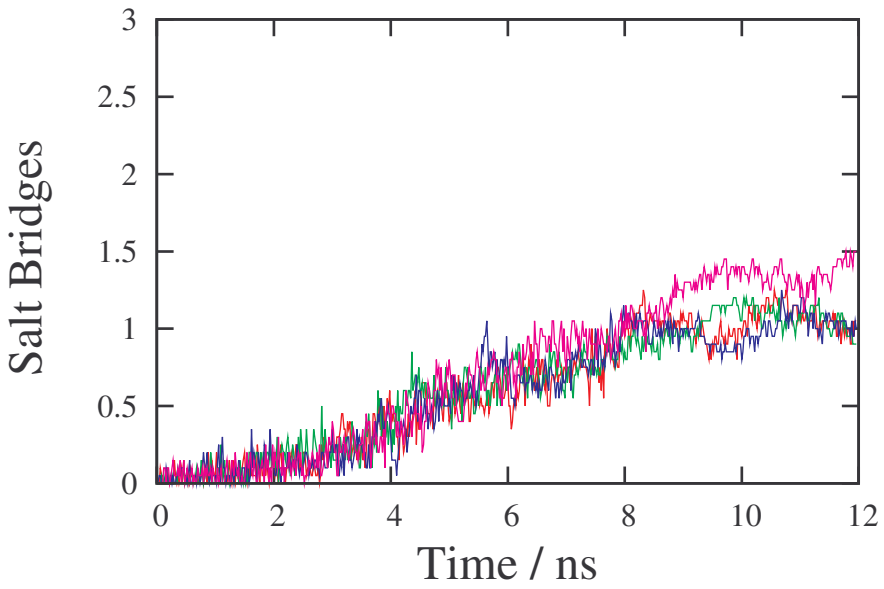


Figure 4.10: Average number of non-native salt bridges across 20 simulation replicas during simulated annealing of AEK17, with no swarm potential present (red) and with non-periodic swarm potentials of strength $A = -25.0$ kcal/mol (green), $A = -50.0$ kcal/mol (blue) and $A = -100.0$ kcal/mol (pink).

$A = -50.0$ kcal/mol greatly increased the performance of the simulated annealing simulations. The swarm average helicity observed during the final 2 ns of simulation increased from the 30% seen in the independent replica simulations to 63% for $A = -25.0$ and 85% for $A = -50.0$ kcal/mol (Figure 4.11). The stronger $A = -100.0$ kcal/mol potential had a negative effect on the helicity observed during the simulation, with the swarm average falling to 10%.

The presence of periodic swarm potentials in the dynamics of the simulated annealing trajectories increased the number of native salt bridges of observed during the latter stages of the simulations. The average of 1.5 salt bridges seen during the final 2 ns of the independent simulations increased to 2.2, 2.7 and 2.2 native salt bridges respectively for potentials of strengths $A = -25.0$, $A = -50.0$ and $A = -100.0$ kcal/mol. The number of non-native salt bridges detected also

Table 4.5: Final average helicity and native salt bridge occupancy for simulated annealing of AEK17 with non-periodic potentials.

	$A = 0.0$[a]		$A = -25.0$[a]		$A = -50.0$[a]		$A = -100.0$[a]	
Replica	Helicity[b]	SB[c]	Helicity[b]	SB[c]	Helicity[b]	SB[c]	Helicity[b]	SB[c]
1	0.28	1.9	0.67	0.9	0.02	0.0	0.39	2.7
2	0.10	2.0	0.36	0.8	0.53	0.1	0.29	0.3
3	0.07	2.1	0.30	1.0	0.42	1.5	0.16	1.0
4	0.70	2.7	0.72	2.5	0.65	2.5	0.75	2.6
5	0.58	1.0	0.09	0.0	0.38	0.2	0.48	2.8
6	0.37	0.0	0.36	2.8	0.64	2.6	0.12	0.0
7	0.16	0.0	0.40	2.2	0.20	1.0	0.19	0.0
8	0.24	2.0	0.25	2.7	0.48	1.0	0.39	2.0
9	0.00	0.9	0.62	1.5	0.29	0.5	0.53	1.1
10	0.53	1.2	0.26	1.0	0.42	0.9	0.00	0.0
11	0.28	2.9	0.21	2.8	0.30	1.0	0.37	1.9
12	0.06	1.9	0.09	2.0	0.44	1.0	0.34	1.1
13	0.38	1.7	0.24	0.3	0.58	0.8	0.02	0.4
14	0.19	2.2	0.43	1.7	0.08	1.5	0.41	1.9
15	0.25	1.7	0.75	2.2	0.50	1.2	0.14	1.9
16	0.00	0.0	0.31	1.2	0.62	2.6	0.11	0.0
17	0.44	1.8	0.12	1.5	0.23	1.0	0.63	0.7
18	0.81	2.3	0.07	1.0	0.17	0.7	0.29	1.9
19	0.40	1.0	0.73	2.5	0.39	1.8	0.32	1.0
20	0.09	0.0	0.25	2.0	0.21	0.0	0.71	2.6

[a] Swarm potential strength (kcal/mol).
[b] Average fractional helicity.
[c] Average number of native salt bridges.

decreased relative to the independent replicas, falling from 1.1 to 0.3 for the $A = -25.0$ kcal/mol potential, and less than 0.1 for both the $A = -50.0$ and $A = -100.0$ kcal/mol potentials (Table 4.5).

In terms of the individual simulation replicas, 12 trajectories in the $A = -25.0$ kcal/mol simulation folded to structures with a greater than 70% average helicity and average salt bridge occupancy above 2.0 over the final 2 ns, while 17 of the $A = -50.0$ kcal/mol trajectories did (Table 4.6). This is a significant improvement over the single independent replica that achieved a fully folded conformation. In contrast, the stronger $A = -100.0$ kcal/mol potential had an adverse

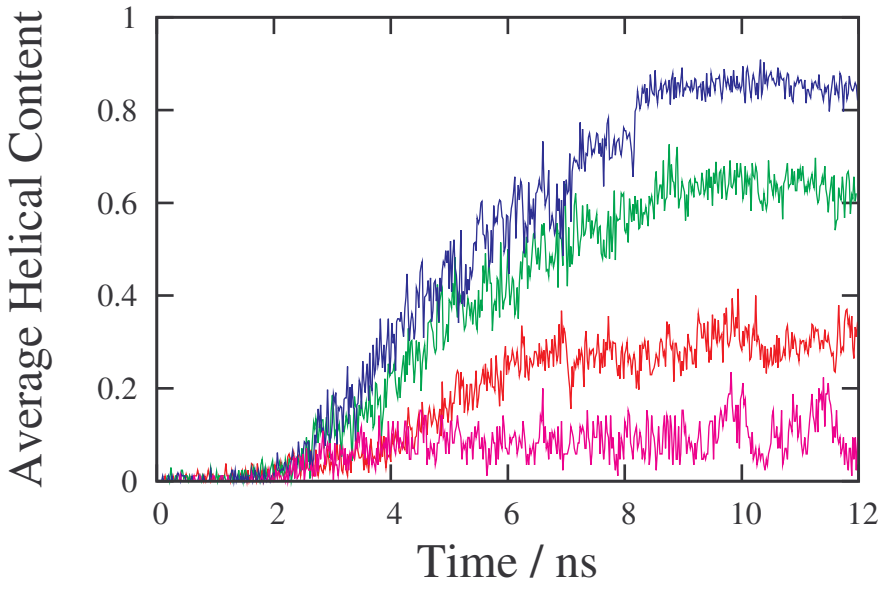


Figure 4.11: Average fractional helicity across 20 simulation replicas during simulated annealing of AEK17, with no swarm potential present (red) and with periodic swarm potentials of strength $A = -25.0$ kcal/mol (green), $A = -50.0$ kcal/mol (blue) and $A = -100.0$ kcal/mol (pink).

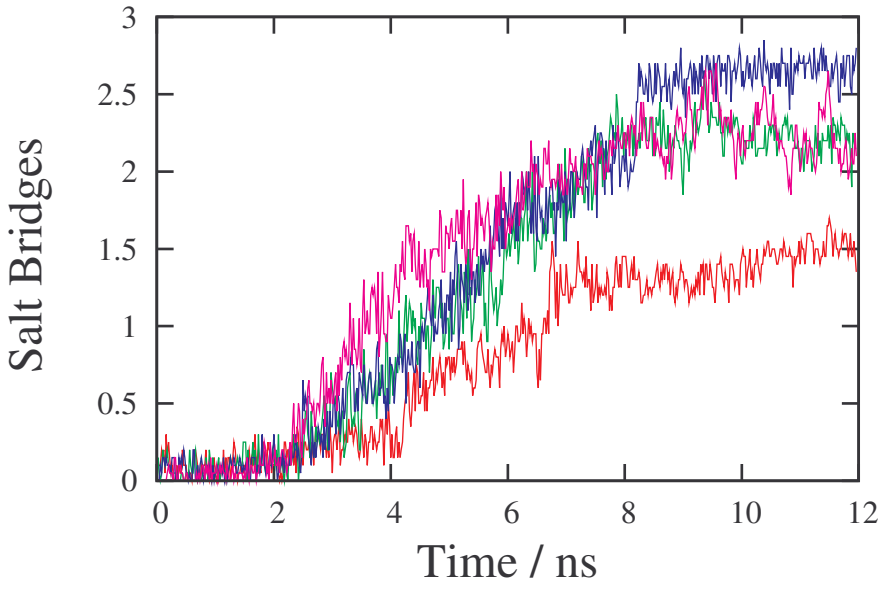


Figure 4.12: Average number of native salt bridges across 20 simulation replicas during simulated annealing of AEK17, with no swarm potential present (red) and with periodic swarm potentials of strength $A = -25.0$ kcal/mol (green), $A = -50.0$ kcal/mol (blue) and $A = -100.0$ kcal/mol (pink).

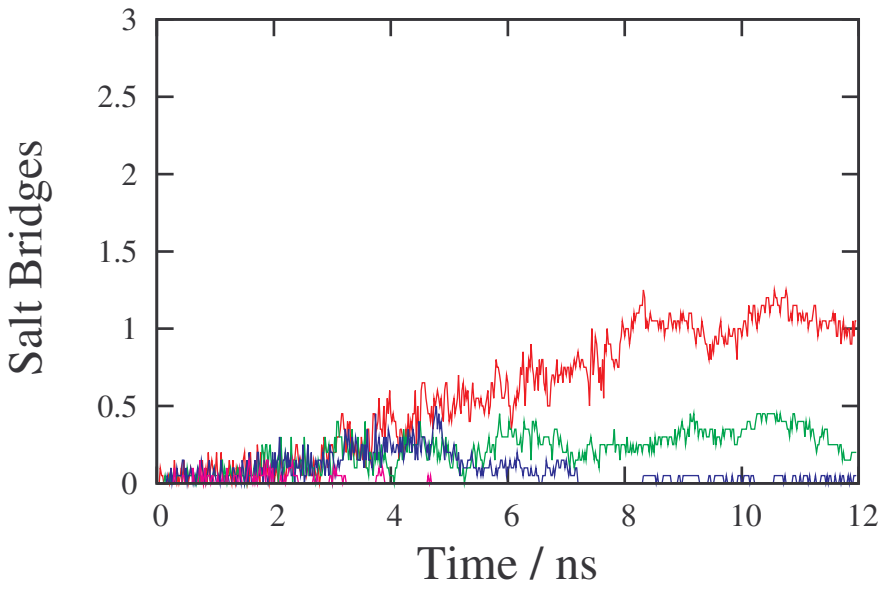


Figure 4.13: Average number of native salt bridges across 20 simulation replicas during simulated annealing of AEK17, with no swarm potential present (red) and with periodic swarm potentials of strength $A = -25.0$ kcal/mol (green), $A = -50.0$ kcal/mol (blue) and $A = -100.0$ kcal/mol (pink).

effect on folding performance, with no trajectories folding completely.

4.3.3.3 Trp-cage

4.3.3.3.1 Annealing Schedule A Following the 25 ns annealing schedule A (Table 4.3), two of the independent replicas of Trp-cage (Replicas 4 and 18) folded to structures that displayed an average backbone RMSD of less than 2.0 Å, relative to the NMR derived reference native structure, over the final 2 ns of simulation (Figure 4.14). Two further replicas have a backbone RMSD lower than 2.5 Å (Replicas 1 and 2). With the swarm potential present in the system, two simulation replicas (Replicas 8 and 16) folded to structures with an backbone RMSD lower than 2.0 Å (Figure 4.15), with 12 others (Replicas 1, 4–6, 8, 9, 11-13, 18–20) folding to structures with an RMSD lower than 2.5 Å. Overall the average

Table 4.6: Final average helicity and native salt bridge occupancy for simulated annealing of AEK17 with periodic potentials.

	$A = 0.0$[a]		$A = -25.0$[a]		$A = -50.0$[a]		$A = -100.0$[a]	
Replica	Helicity[b]	SB[c]	Helicity[b]	SB[c]	Helicity[b]	SB[c]	Helicity[b]	SB[c]
1	0.28	1.9	0.76	2.6	0.91	2.8	0.09	2.1
2	0.10	2.0	0.72	2.5	0.92	2.8	0.12	2.3
3	0.7	2.1	0.47	2.8	0.92	2.9	0.13	2.3
4	0.69	2.7	0.14	1.0	0.30	1.5	0.11	2.2
5	0.58	1.0	0.48	1.7	0.91	2.8	0.11	2.1
6	0.37	0.0	0.79	2.8	0.91	2.8	0.11	2.3
7	0.16	0.0	0.80	2.7	0.69	2.2	0.09	2.2
8	0.24	2.0	0.74	2.7	0.91	2.8	0.13	2.2
9	0.00	0.9	0.81	2.8	0.87	2.8	0.11	2.3
10	0.53	1.2	0.80	2.6	0.92	2.8	0.08	2.2
11	0.28	2.9	0.72	1.5	0.91	2.8	0.08	2.2
12	0.06	1.9	0.00	0.0	0.91	2.8	0.12	2.2
13	0.38	1.7	0.59	2.5	0.91	2.8	0.10	2.1
14	0.19	2.2	0.80	2.7	0.91	2.8	0.12	2.4
15	0.25	1.7	0.57	2.2	0.91	2.8	0.09	2.2
16	0.00	0.0	0.80	2.5	0.92	2.8	0.11	2.2
17	0.44	1.8	0.42	1.7	0.91	2.8	0.09	2.3
18	0.81	2.3	0.77	2.7	0.84	2.8	0.08	2.1
19	0.40	1.0	0.74	1.0	0.91	2.8	0.10	2.2
20	0.09	0.0	0.77	2.7	0.66	1.6	0.10	2.1

[a] Swarm potential strength (kcal/mol).
[b] Average fractional helicity.
[c] Average number of native salt bridges.

backbone RMSD of the 20 independent replicas was 3.47 Å, compared to 2.93 Å for the swarm directed replicas.

The independent replica with the lowest backbone RMSD (Replica 18: backbone RMSD 1.72 Å) also has the lowest heavy atom RMSD of 2.78 Å (Figure 4.16). This is the only independent replica that folded to a structure with a heavy atom RMSD of less than 3.0 Å. The two swarm-directed replicas which displayed the lowest backbone RMSDs (Replica 16: backbone RMSD 1.29 Å and Replica 8: backbone RMSD 1.31 Å) also displayed the lowest heavy atom RMSDs of 2.22 and 2.23 Å respectively (Figure 4.17). Again the overall average RMSD

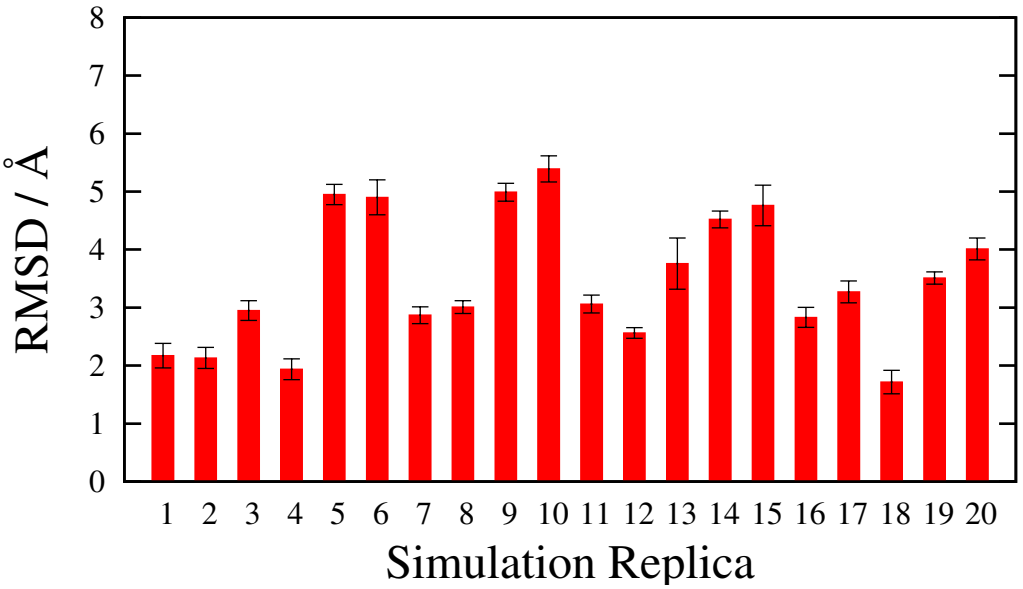


Figure 4.14: Average backbone RMSD of Trp-cage, over final 2 ns of simulation, for 20 independent replicas (Annealing Schedule A).

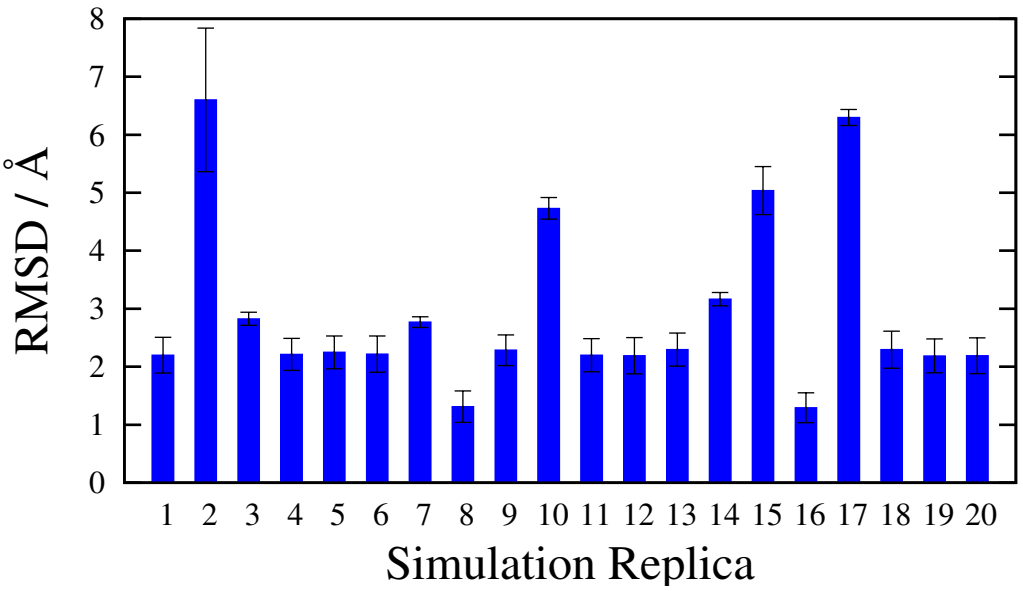


Figure 4.15: Average backbone RMSD of Trp-cage, over final 2 ns of simulation, for 20 swarm directed replicas (Annealing Schedule A).

was higher for the 20 independent replicas (4.90 Å) than it was for the swarm simulations (4.41 Å).

The largest deviations in backbone RMSD occur in the terminal ends of the chain. In particular the N-terminal α-helix is not fully folded, and a non-native salt bridge between the terminal amino and carboxyl groups is present, in many of both the independent and swarm directed replicas (Figure 4.18). Removing the three N-terminal residues Asn20, Leu21 and Tyr22 and the C-terminal Ser39 residue from the analysis, seven independent replicas (Replicas 1, 2, 4, 7, 13, 16 and 18) fold to structures with an RMSD relative to the NMR structure of less than 2.0 Å, with one of these (Replica 18) folding to an RMSD below 1.5 Å (Figure 4.19). The average backbone RMSD across the 20 replicas was 2.76 Å. In the swarm potential directed simulations, 12 replicas folded to a backbone RMSD of less than 1.5 Å (Replicas 1, 4–6, 8, 9, 11, 12, 16, 18–20), with three others (Replicas 3, 7 and 13) folding to below 2.0 Å (Figure 4.20). The average backbone RMSD across the swarm was 1.90 Å.

The independent replica which shows the lowest backbone RMSD for residues 23 to 38 (Replica 18: backbone RMSD of residues 23 to 80 1.19 Å) also shows the lowest heavy atom RMSD for these residues, with a value of 1.93 Å (Figure 4.21). This is the only independent replica that folds to a structure that agrees with the NMR structure to within 2.0 Å for these residues. Two others (Replicas 2 and 13) agree to within 2.5 Å. The three other replicas that showed a backbone RMSD, for residues 23 to 38, of below 2.0 Å, relative to the NMR structure, had heavy atom RMSDs in the range 2.8 – 3.6 Å. The poor heavy atom agreement for these replicas is due to an incorrect orientation of the Trp25 residue (Figure 4.23). The average heavy atom RMSD for residues 23 to 38, across the 20 independent replicas was 4.15 Å.

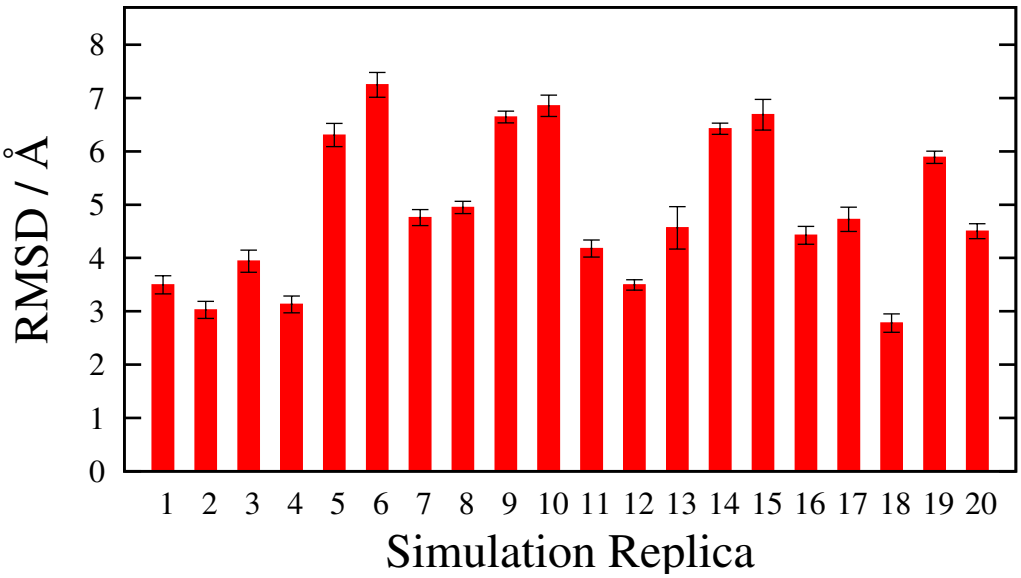


Figure 4.16: Average heavy-atom RMSD of Trp-cage, over final 2 ns of simulation, for 20 independent replicas (Annealing Schedule A).

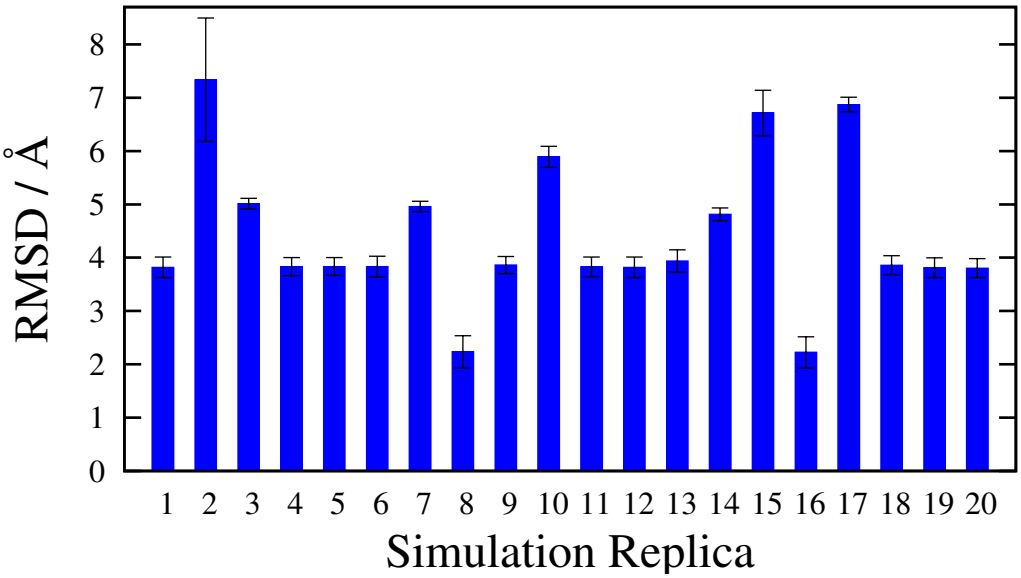
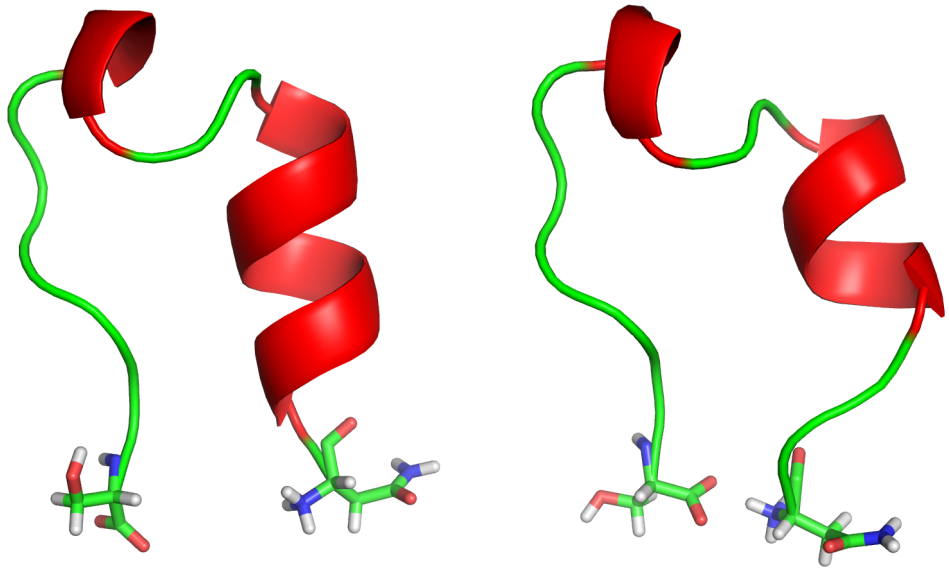


Figure 4.17: Average heavy-atom RMSD of Trp-cage, over final 2 ns of simulation, for 20 swarm directed replicas (Annealing Schedule A).



(a) Native NMR structure **(b)** Folded Replica

Figure 4.18: Comparison of native NMR structure with a misfolded replica showing incomplete helix formation and non-native salt bridge.

All 12 of the swarm directed replicas that folded to within 1.5 Å of the native structure of residues 23 to 38 showed a heavy atom RMSD for these residues of below 2.2 Å (Figure 4.22). Replica 8, which had the lowest backbone RMSD of 1.12 Å, also had the lowest heavy atom RMSD of 1.97 Å. This was the only replica that folded to a structure with a heavy atom RMSD below 2.0 Å. The three replicas that folded to structures with backbone RMSDs between 1.5 and 2.0 showed heavy atom RMSDs in the range 2.5 – 3.7 Å. Again the poor heavy atom agreement for these replicas was due to an incorrect orientation of the Trp25 side chain. The average heavy atom RMSD across the 20 swarm members was lower than for the independent replicas, falling to 2.99 Å.

4.3.3.3.2 Annealing Schedule B The average backbone RMSD across 20 independent replicas, following the longer simulated annealing schedule B (Table

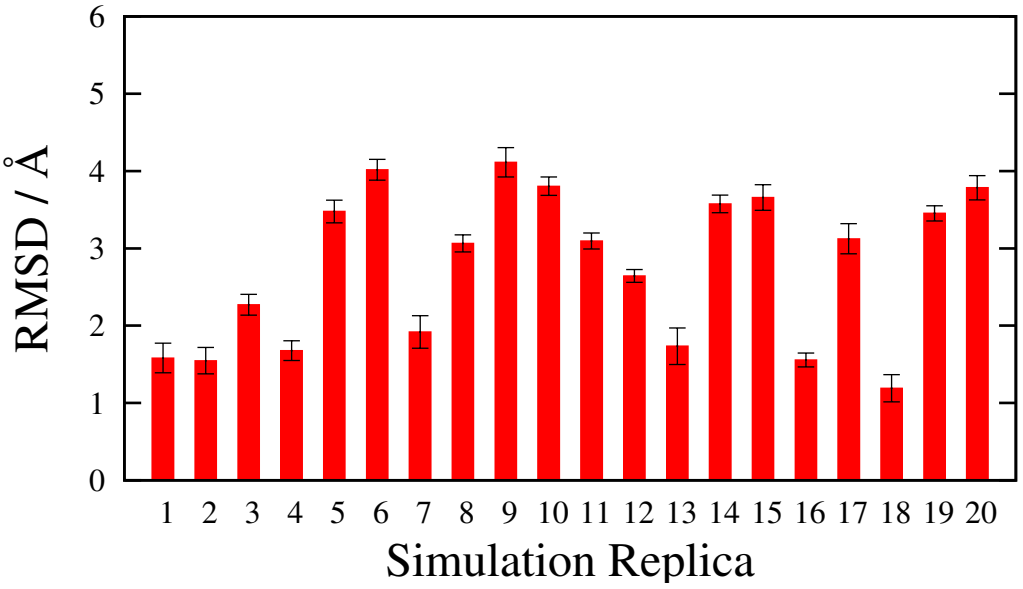


Figure 4.19: Average backbone RMSD of residues 23 to 38 of Trp-cage, over final 2 ns of simulation, for 20 independent replicas (Annealing Schedule A).

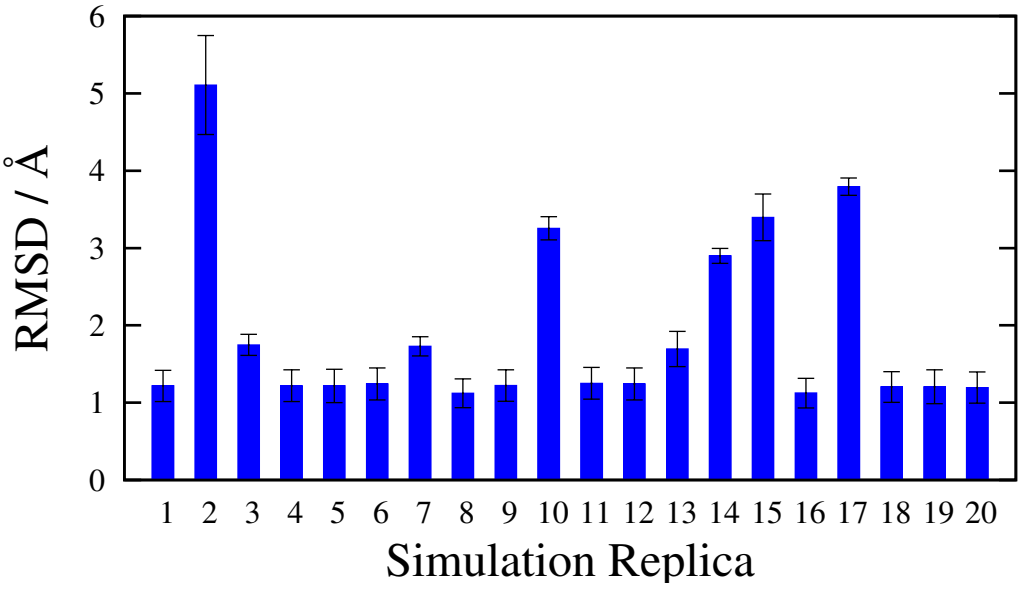


Figure 4.20: Average backbone RMSD of residues 23 to 38 of Trp-cage, over final 2 ns of simulation, for 20 swarm directed replicas (Annealing Schedule A).

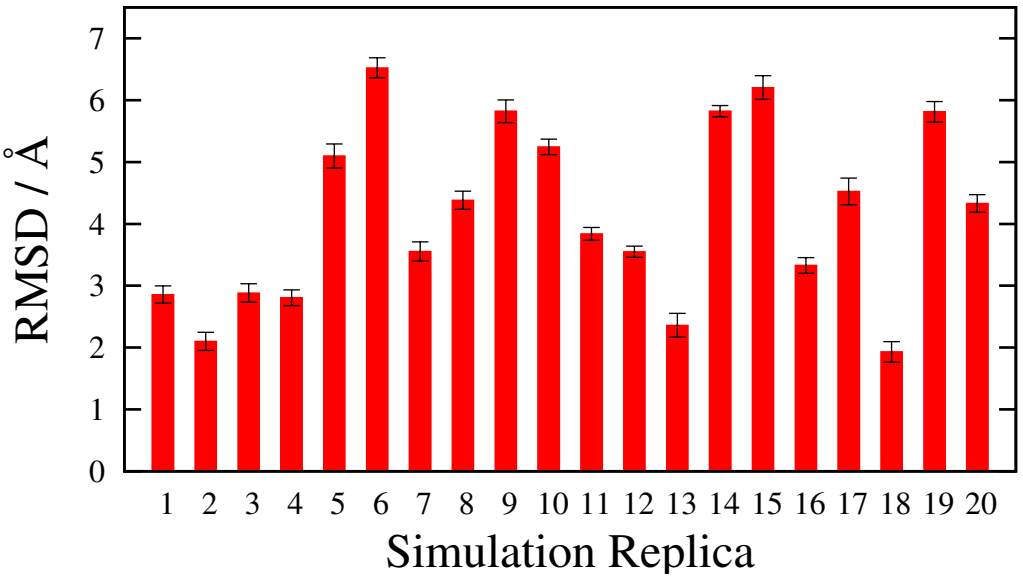


Figure 4.21: Average heavy atom RMSD of residues 23 to 38 of Trp-cage, over final 2 ns of simulation, for 20 independent replicas (Annealing Schedule A).

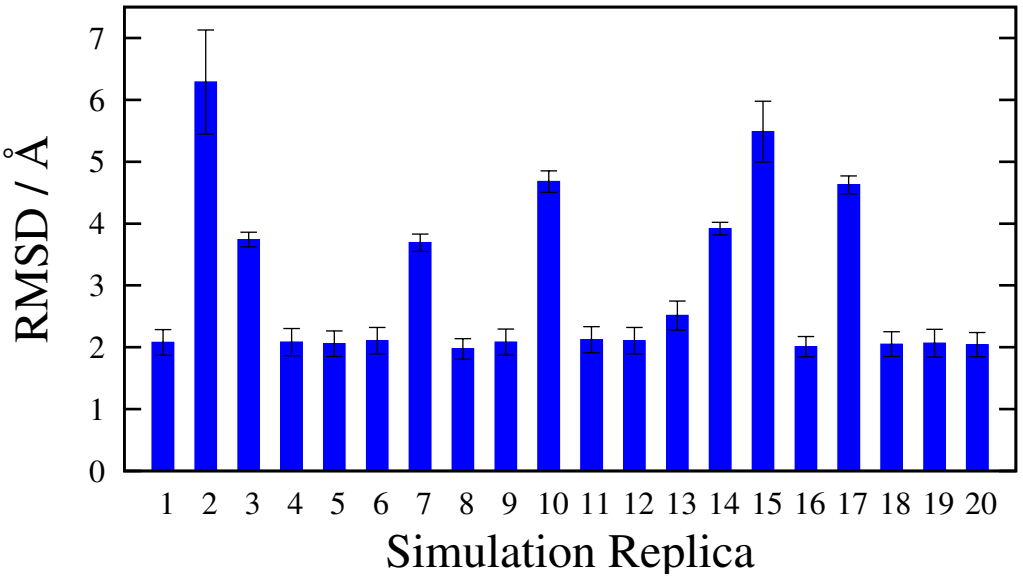
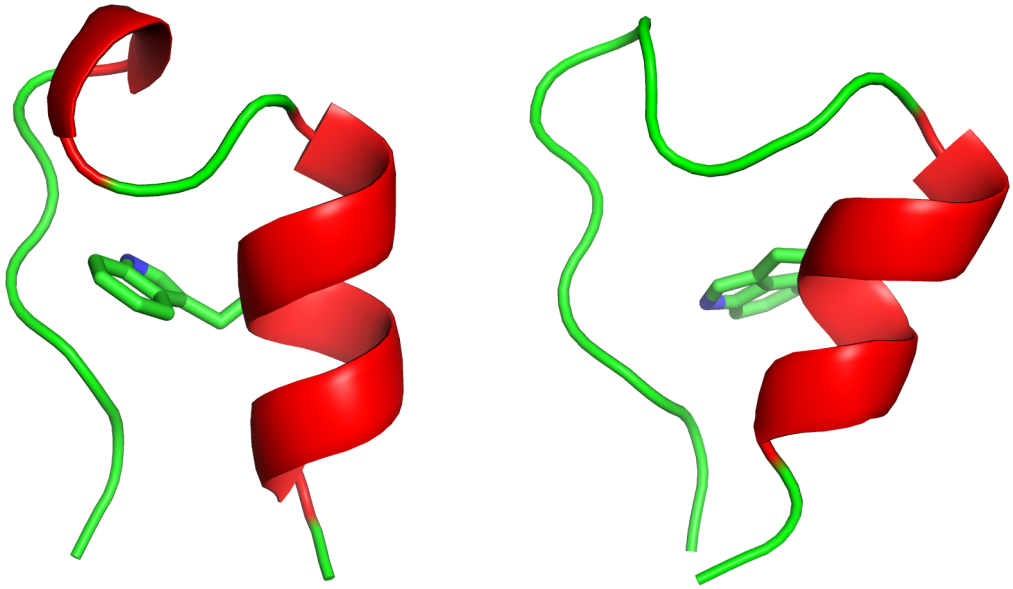


Figure 4.22: Average heavy atom RMSD of residues 23 to 38 of Trp-cage, over final 2 ns of simulation, for 20 swarm directed replicas (Annealing Schedule A).



(a) Native NMR structure **(b)** Misfolded replica

Figure 4.23: Comparison of native NMR structure with a misfolded replica showing incorrect Trp25 orientation.

4.4), was 3.25 Å. Four of the replicas displayed an RMSD of lower than 2.0 Å(Replicas 1, 3, 4 and 9), with four others (Replicas 2, 7, 8 and 12) having an RMSD below 2.5 Å (Figure 4.24). Replicas 1 and 4 showed the best agreement with the reference NMR structure, both having an RMSD of 1.87 Å.

With the swarm potential present the average backbone RMSD of across the swarm members was reduced significantly, dropping to 1.63 Å. 16 of the swarm members folded to structures with an RMSD of lower than 1.5 Å (Replicas 1–4, 6–10, 13–15 and 17–20), with the lowest, Replica 10, having an RMSD of 1.25 Å (Figure 4.25). One other replica (Replica 16) had an RMSD of lower than 2.0 Å, and one (Replica 12) had an RMSD below 2.5 Å.

The swarm potential also significantly improved the agreement in heavy atom RMSD between the simulation replicas and the reference NMR structure. The

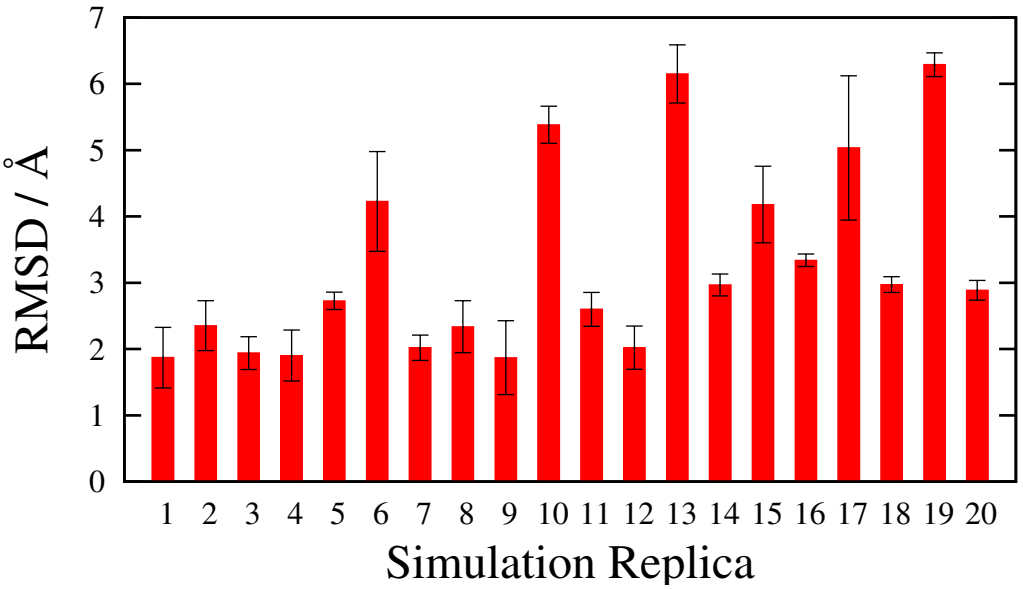


Figure 4.24: Average backbone RMSD of Trp-cage, over final 2 ns of simulation, for 20 independent replicas (Annealing Schedule B).

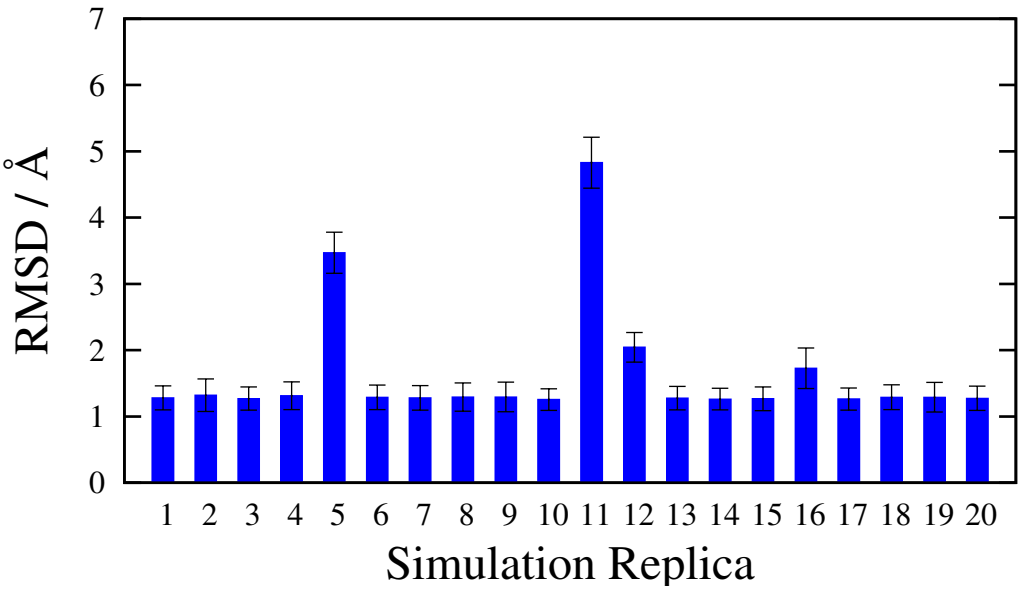


Figure 4.25: Average backbone RMSD of Trp-cage, over final 2 ns of simulation, for 20 swarm directed replicas (Annealing Schedule B).

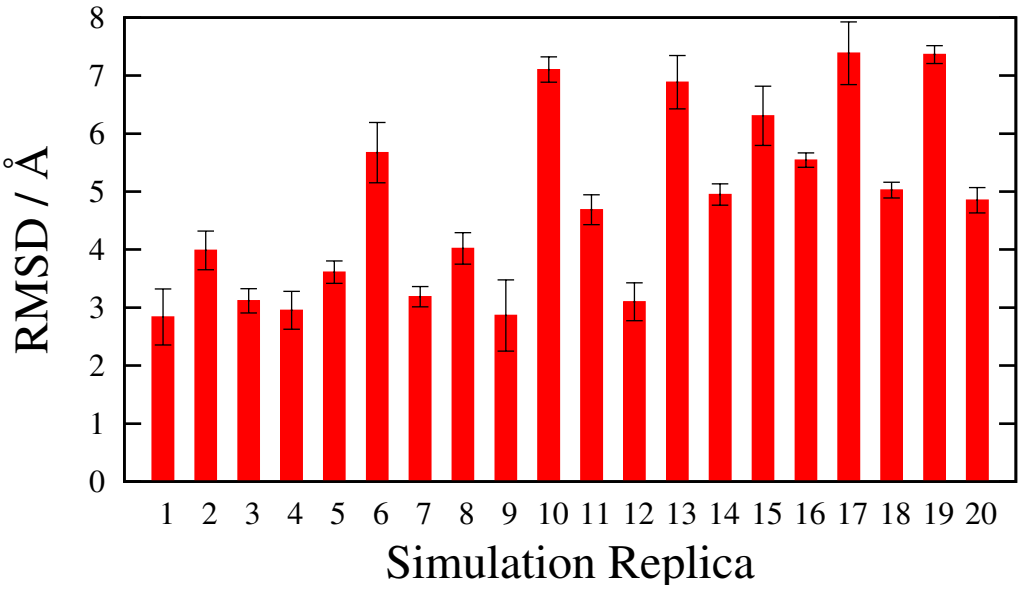


Figure 4.26: Average heavy-atom RMSD of Trp-cage, over final 2 ns of simulation, for 20 independent replicas (Annealing Schedule B).

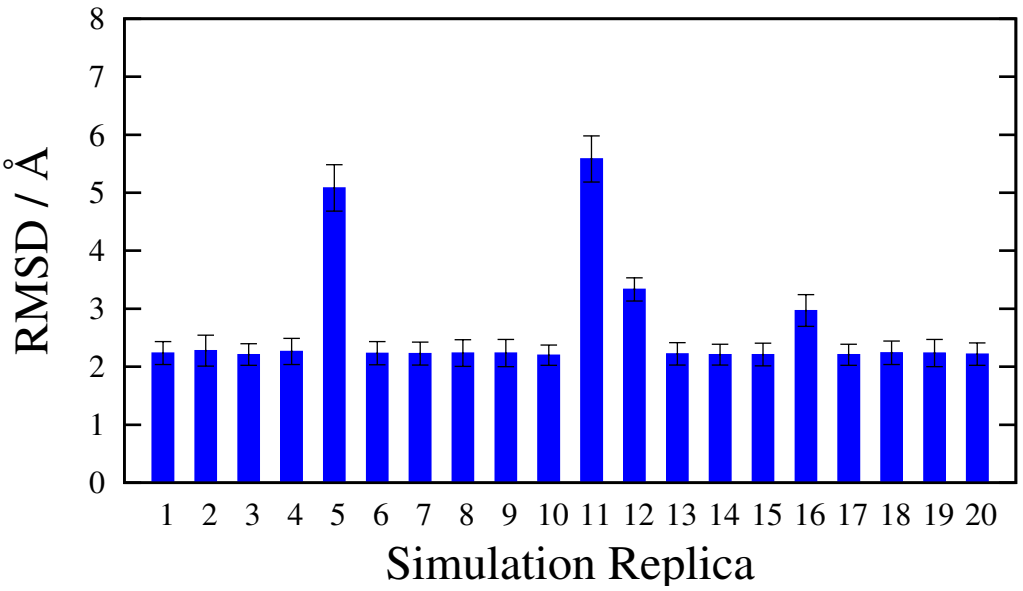


Figure 4.27: Average heavy-atom RMSD of Trp-cage, over final 2 ns of simulation, for 20 swarm directed replicas (Annealing Schedule B).

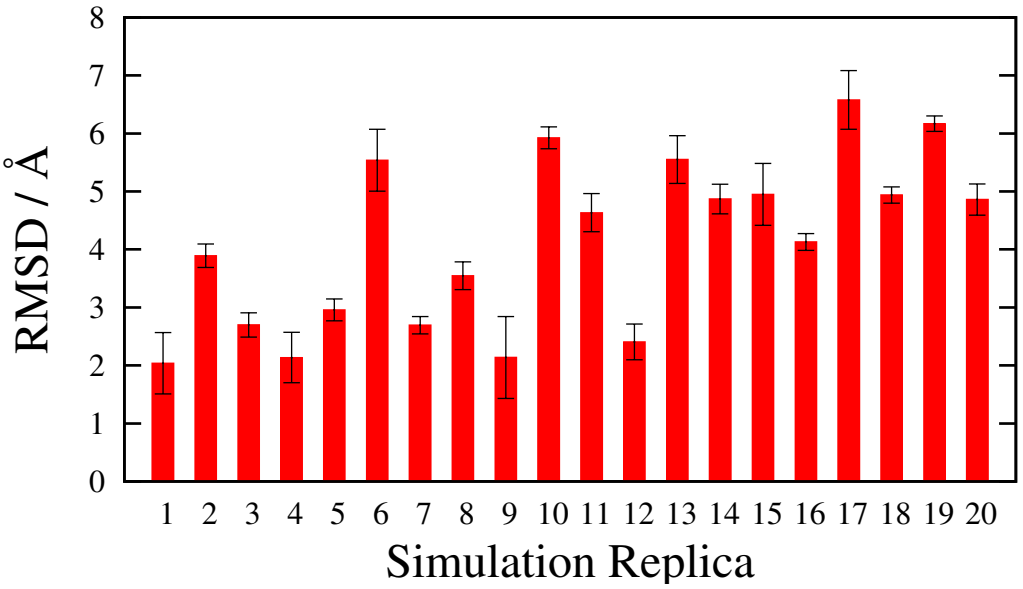


Figure 4.28: Average RMSD of Trp-cage backbone atoms and side chains of residues that comprise the cage, over final 2 ns of simulation, for 20 independent replicas (Annealing Schedule B).

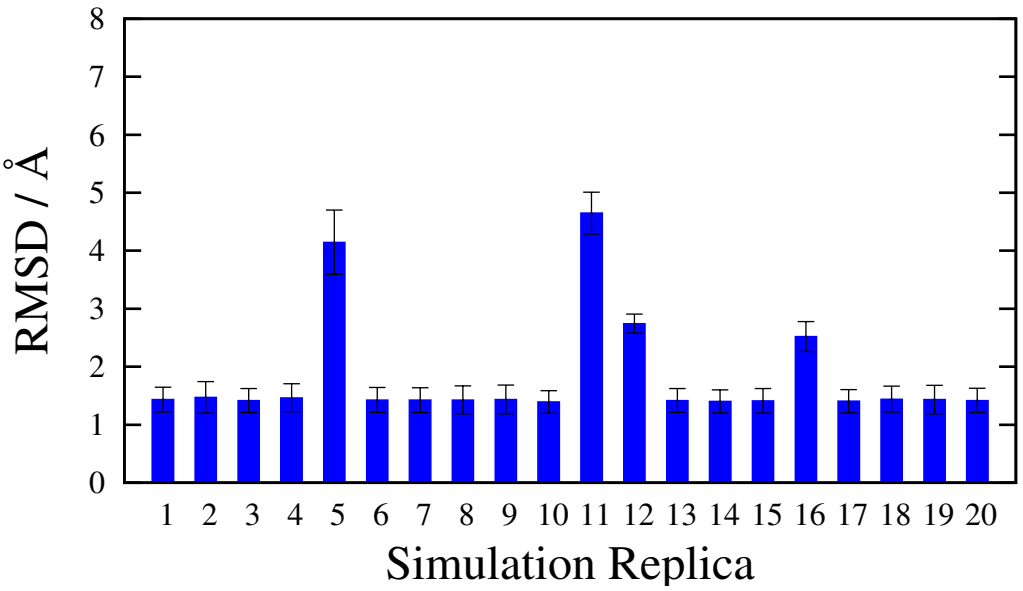


Figure 4.29: Average RMSD of Trp-cage backbone atoms and side chains of residues that comprise the cage, over final 2 ns of simulation, for 20 swarm directed replicas (Annealing Schedule B).

independent replicas folded to structures that had an average heavy atom RMSD, across the 20 replicas, of 4.77 Å. The average across the swarm directed replicas was 2.63 Å. Individually, three independent replicas (Replicas 1, 4 and 9) folded to a heavy atom RMSD below 3.0 Å, with three others (Replicas 3, 7 and 12) folding to within 3.5 Å (Figure 4.26). In comparison, 16 swarm directed replicas folded to structures with a heavy atom RMSD of below 2.5 Å (Replicas 1–4, 6–10, 13–15 and 17–20), with one other (Replica 16) folding to within 3.0 Å and another (Replica 12) folding to below 3.5 Å (Figure 4.27).

The most highly conserved side chain positions in the NMR derived ensemble of structures are those that comprise the cage structure, namely Tyr22, Trp25 and proline residues 31 and 36 – 38. Limiting the RMSD analysis to side chain heavy atoms of these residues, and the backbone of the whole chain, the independent replicas displayed an RMSD, across the 20 simulations, of 4.13 Å, while the swarm directed replicas folded to an average of 1.62 Å. Four independent replicas (Replicas 1, 4, 9 and 12) display an RMSD of below 2.5 Å, while three other (Replicas 3, 5 and 7) have an RMSD below 3.0 Å (Figure 4.28). The 16 swarm replicas that have a heavy atom RMSD of below 2.5 all show an RMSD of below 1.5 Å for their backbone atoms and heavy cage residue atoms, while the replicas 16 and 12, which display heavy atom RMSDs in the range 2.5 –3.5, have RMSDs of 2.52 and 2.74 Å respectively.

The 20 independent replica folded to structures that, on average, violated 35% of the 168 NOE restraints used to generate the reference structure of Trp-cage (Figure 4.30). The lowest violations were shown by replicas 1 and 4, both replicas violating 20% of NOE restraints. Two other replicas (Replicas 9 and 12) violated 22% of NOE restraints. All other replicas violated greater than 30% of NOE restraints. Overall the fractional violation of the swarm simulations was lower,

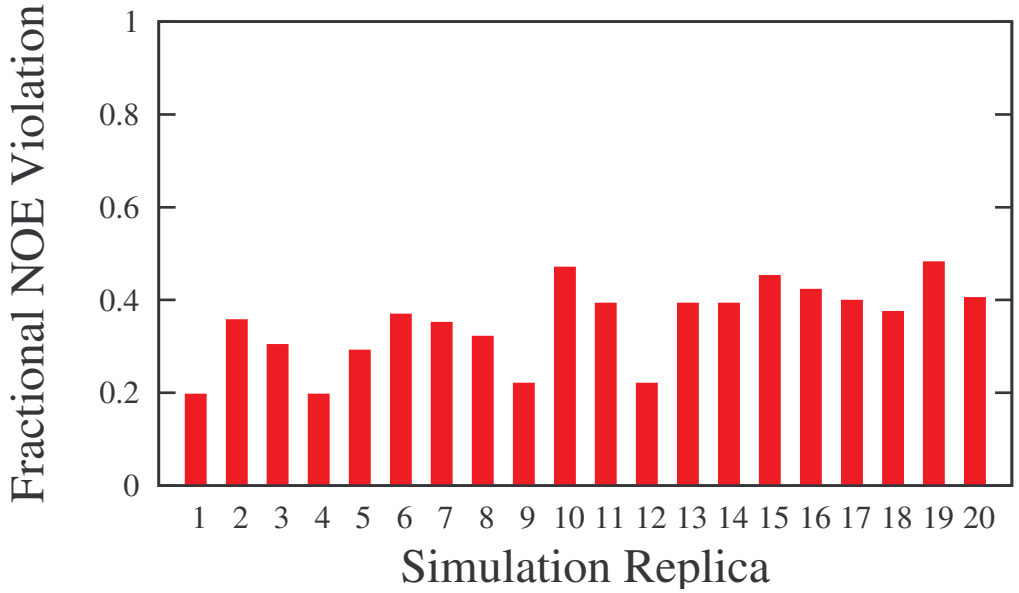


Figure 4.30: Average NOE violation, over final 2 ns of 40 ns simulation, for 20 independent replicas (Annealing Schedule B).

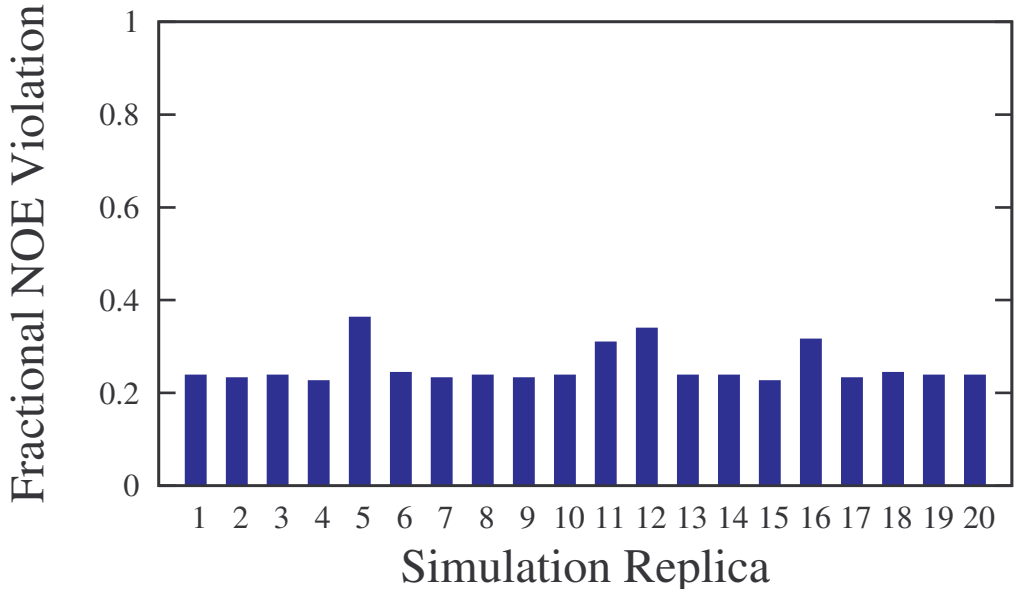


Figure 4.31: Average NOE violation, over final 2 ns of 40 ns simulation, for 20 swarm directed replicas (Annealing Schedule B).

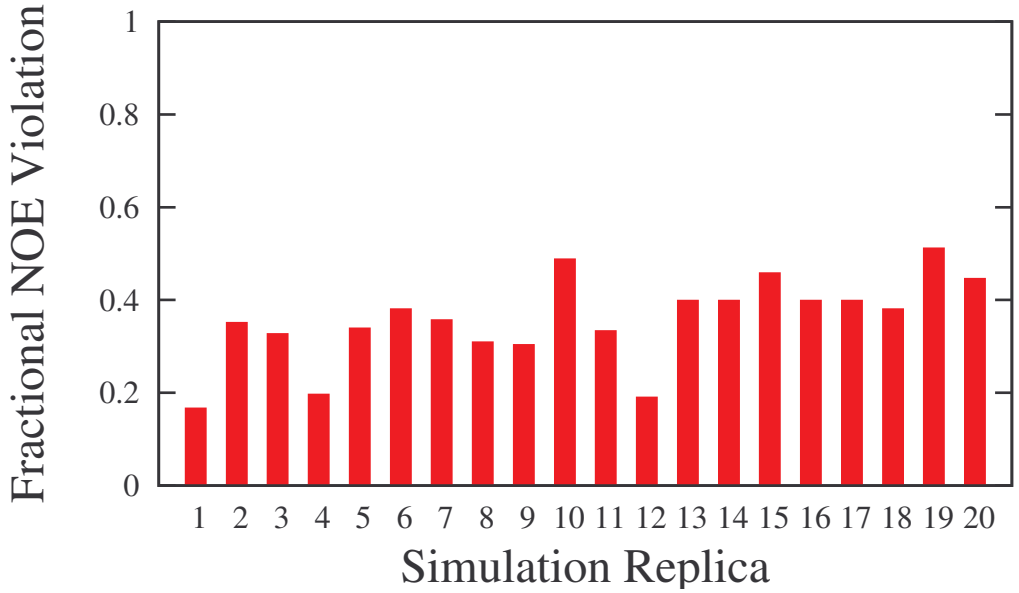


Figure 4.32: Average NOE violation, over final 2 ns of 45 ns simulation, for 20 independent replicas (Annealing Schedule B).

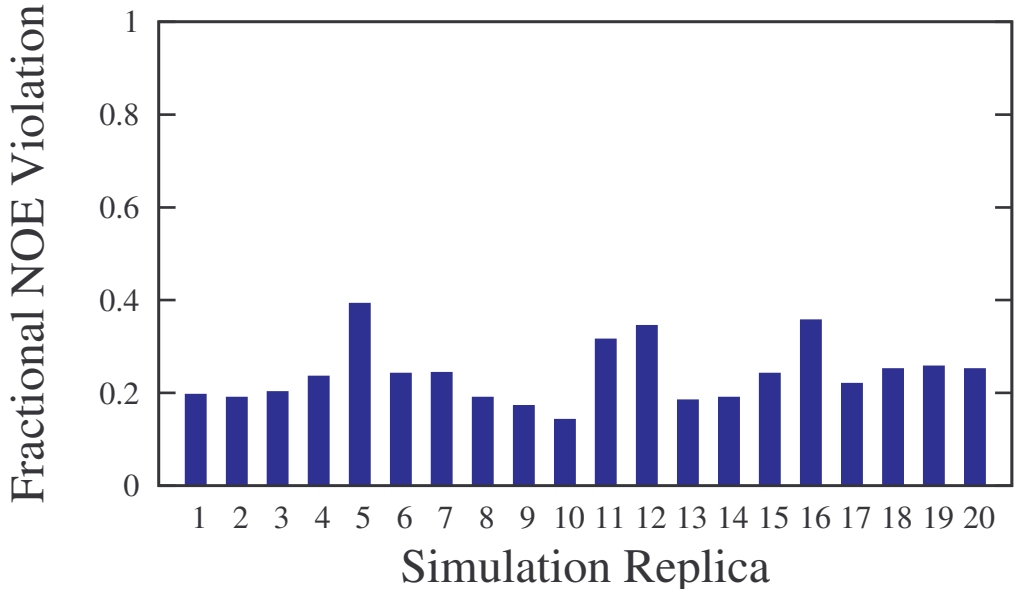


Figure 4.33: Average NOE violation, over final 2 ns of 45 ns simulation, for 20 swarm directed replicas (Annealing Schedule B).

with an average of 26% NOE restraints violated (Figure 4.31). Individually, the swarm replicas which showed the lowest violations displayed fractional violations higher than the 20% achieved by the best folded individual replicas. The swarm directed replicas with the lowest violations (Replicas 2, 5, 7, 9, 15 and 17) violated 23% of NOE restraints, while 10 other swarm members (Replicas 1, 3, 6, 8, 10, 13, 14, 18–20) violated 24% of restraints. The four remaining replicas all violated greater than 30% of NOE restraints.

Following relaxation of the swarm potential, as described in section 4.3.2.3, the average NOE violation of the individual replicas rose slightly to 36%, while the average NOE violation of the swarm replicas fell to 24% (Figures 4.32 and 4.33). Three of the independent replicas (Replicas 1, 4 and 12) now violated less than 20% of NOE restrainsts, with the the lowest violating 17%. The rest of the replicas all violated over 30% of restraints. Seven of the swarm directed replicas violated less than 20% of NOE restraints (Replicas 1, 2, 8–10, 13–14), with the lowest value being 14%. Nine other replicas violated less than 30% of restraints.

4.3.4 Discussion and Conclusions

4.3.4.1 Relative Performance of Non-periodic and Periodic Swarm Potentials

4.3.4.1.1 Polyalanine As outlined in the results above, in the absence of any swarm forces, a number of the independent replicas of polyalanine become trapped in non-helical conformations, with only five replicas folding correctly. The misfolded replicas display a variety of partially folded α-helices, 3_{10}-helices and random coil structures (Figure 4.34). The high proportion of incorrectly folded

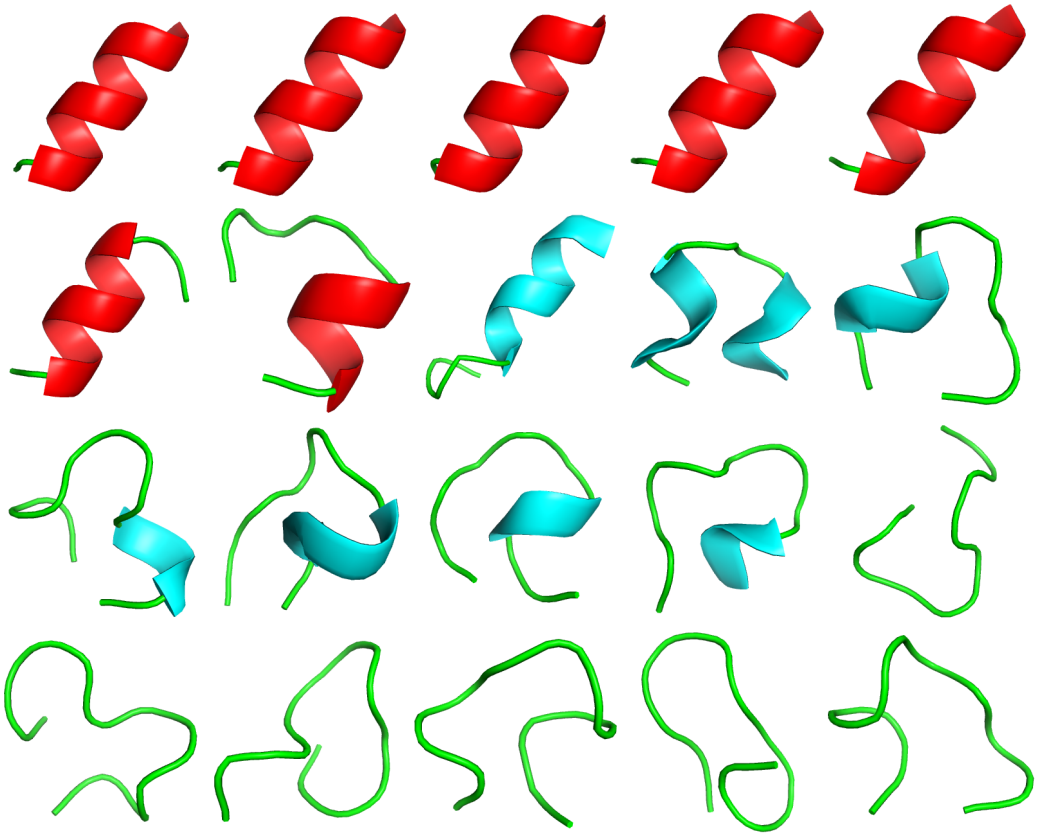


Figure 4.34: Final folded conformations, following simulated annealing, for 20 independent replicas of polyalanine (Red: α-helix; Blue: 3_{10}-helix; Green: random coil).

structures obtained from these unbiased simulations suggest that the 1.2 ns annealing schedule employed here does not have a sufficiently slow rate of cooling to prevent the polyalanine replicas from becoming trapped in non-native energy basins. This means longer timescale simulations are required to reliably reach the global minimum of the free energy surface.

The use of a non-periodic dihedral angle definition, as described by Huber and van Gunsteren,[41] did not improve the performance of the simulated annealing protocol. The values used for parameters A and B in Equation 4.1 were chosen to match those employed by Huber and van Gunsteren. In their simulations they used values of $A = -200.0$ kJ/mol and $B = 0.8$ rad^{-1} to perform simulated

annealing simulations on a set of model molecules. The $A = -50.0$ kcal/mol potential strength was chosen as it is approximately equivalent to -200.0 kJ/mol (-200.0 kJ/mol $= -47.8$ kcal/mol) while keeping the unit choice consistent with that used in Amber.[76] The -25.0 and -100.0 kcal/mol potential strengths were chosen to examine the effect that the potential strength has on the simulation performance. As each of these potentials appeared to hinder the formation of helices during the simulations, weaker potentials of $A = -5.0$, -10.0, -15.0 and -20.0 kcal/mol were used to investigate whether these could improve the helical formation. As no improvement relative to the independent replicas was observed, and indeed the $A = -5.0$ and -10.0 kcal/mol potentials appeared to perform similarly to the unbiased simulations, it was assumed that these potentials were too weak to significantly affect the dynamics of the simulations. During all non-periodic simulations, the misfolded structures obtained were similar to those seen in the independent replica simulations, with a mixture of partially folded α-helix, 3_{10}-helix and random coil conformations.

Using the periodic dihedral angle definition described in Section 4.2.1.1, significant increases in helix formation were observed. With a potential of strength $A = -25.0$ kcal/mol, the number of completely folded replicas obtained from simulated annealing fell from the five seen for the independent replicas to four. The total average helicity over all swarm members did increase however, from 29% for the independent replicas to 39% for the swarm simulation replicas. This is due to an increase in the number of partially folded α-helical structures, and a decrease in the number of unfolded random coil conformations (Figure 4.35). With a swarm potential of -50.0 kcal/mol the number of unfolded structures decreased to five, and the average helicity increased to 75%. The five unfolded structures obtained from the simulation (Figure 4.36) also appeared to show a

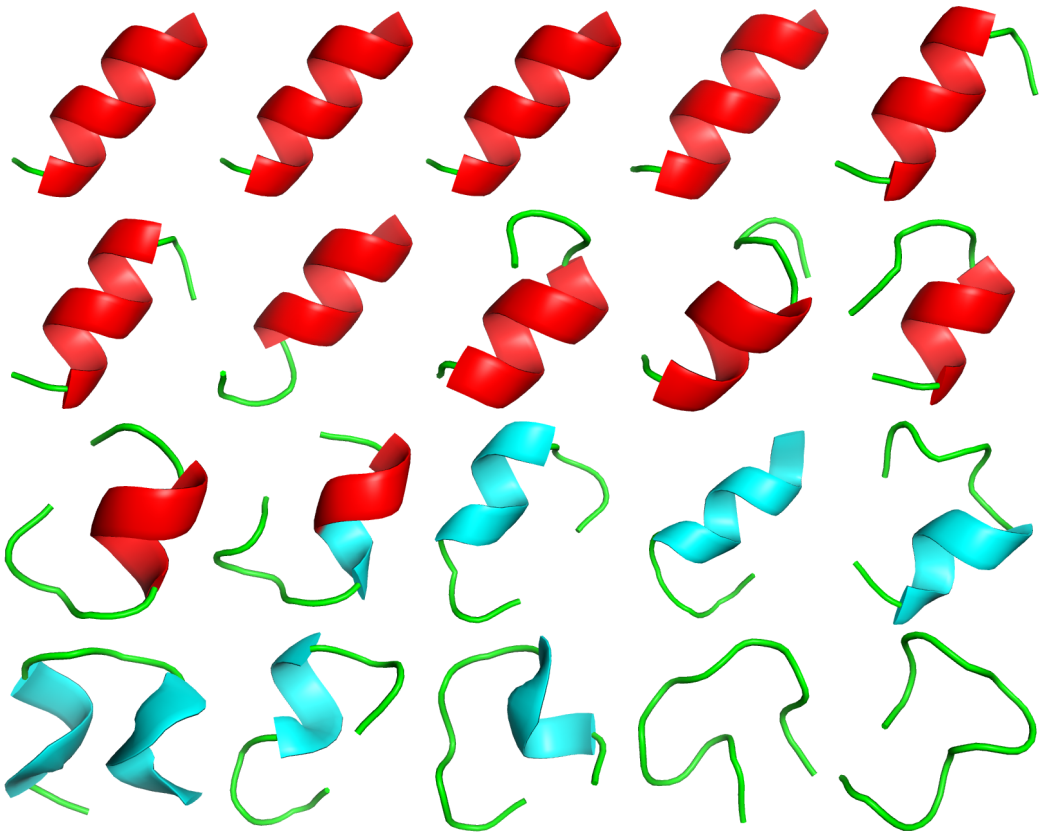


Figure 4.35: Final folded conformations, following simulated annealing, for 20 replicas obtained from a $A = -25.0$ kcal/mol swarm simulation of polyalanine, using a periodic swarm potential (Red: α-helix; Blue: 3_{10}-helix; Green: random coil).

bias towards more helical conformations, with no completely random coil structures being seen.

4.3.4.1.2 AEK17 During simulated annealing of AEK17, only two independent replicas simulation folded to the native structure during the simulation. In a previous study by Wang at el,[114] the folding of AEK17 was investigated via molecular dynamics simulations. They performed 40 folding simulations of length 50 ns. By clustering snapshots obtained from their mid-sections of their simulations, they identified a number of clusters of folding intermediate structures. The most highly populated clusters that they identified were characterised by

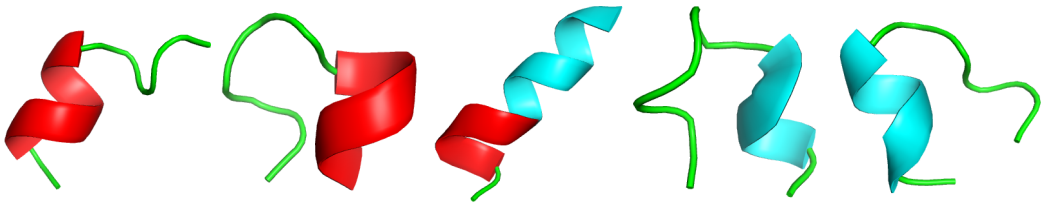


Figure 4.36: Final folded conformations, following simulated annealing, for 5 incorrectly folded replicas obtained from a 20 replica $A = -50.0$ kcal/mol swarm simulation of polyalanine, using a periodic swarm potential (Red: α-helix; Blue: 3_{10}-helix; Green: random coil).

partially folded helix segments separated by L- or U-shaped bends.

For each of the 20 independent replica simulations performed in our work, the average structure over the final 2 ns was calculated, and the simulation snapshot closest to this was identified. Nine of these structures could be characterised as U-shaped 4.37). They show varying degrees of helix formation and native salt bridge occupancies. Five replicas folded to L-shaped structures (Figure 4.38). These, in general, show a greater degree of helicity than the U-shaped structures, but helix formation is frustrated by the presence of non-native salt bridge interactions. In addition to the one correctly folded replica, four others displayed random coil structures, with very little helical content. From these results it can be assumed that slower cooling, and therefore longer simulations, are required to reliably fold AEK17 into its native structure by simulated annealing. The majority of replicas appear to be trapped in intermediate folding structures.

The presence of non-periodic swarm potentials was unable to significantly affect the numbers of folded replicas achieved by the simulations, compared to the unbiased simulations. Of the 20 independent replica simulations, only two folded fully during the 12 ns simulation. While three replicas in the A = −25.0 kcal/mol non-periodic swarm simulation correctly, there was only a small increase in helical

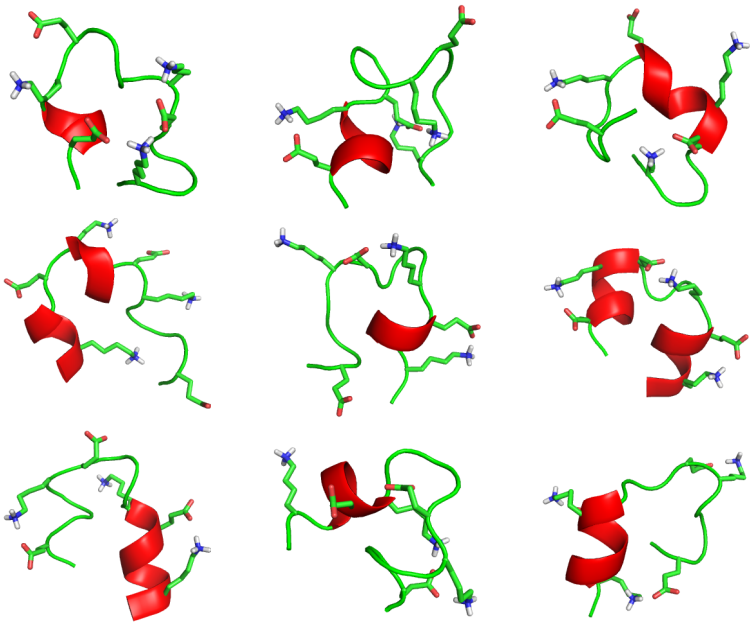


Figure 4.37: Nine U-shaped final structures obtained from 20 independent replica simulated annealing simulations of AEK17.

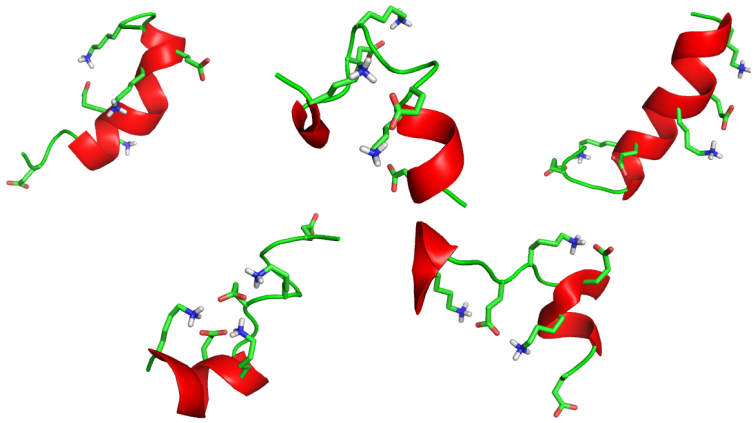


Figure 4.38: Five L-shaped final structures obtained from 20 independent replica simulated annealing simulations of AEK17.

content and native salt bridge formation across all swarm members. In line with the results seen in during the polyalanine simulations, the non-periodic swarm potential either had little effect on the folding performance, or a negative one. Average structures obtained from the latter stages of these simulations showed a similar picture to that seen for the independent replica simulations, with the misfolded replicas characterised by a U- or L-shaped bend, and varying degrees of helical content.

Both the periodic potential simulations with strengths -25.0 and -50.0 kcal/mol showed a marked improvement in helical formation compared to the independent replica simulations, achieving 12 and 17 fully folded replicas respectively. The non-folded replicas from the A $= -25.0$ kcal/mol simulation (Figure 4.39) show a high degree of helicity, with only one displaying no helical content in the average structure, while two of the three non-folded replicas from the -50.0 kcal/mol swarm simulation have almost completely formed helicies (Figure 4.40).

4.3.4.1.3 Summary These results show that by using a swarm potential with a periodic average dihedral definition, it is possible to prevent simulation replicas from becoming trapped in non-native intermediate conformations, and that these potentials significantly outperform non-periodic potentials. It is arguably not surprising that the periodic average dihedral angle definition appears to outperform the non-periodic definition. As the non-periodic average definition ignores the sign of each dihedral, it is unable to discern the difference between an α-helix or a left-handed helix, or a gauche+ or gauche− side chain conformation. This could lead to the swarm potential biasing simulation replicas towards an incorrect structure.

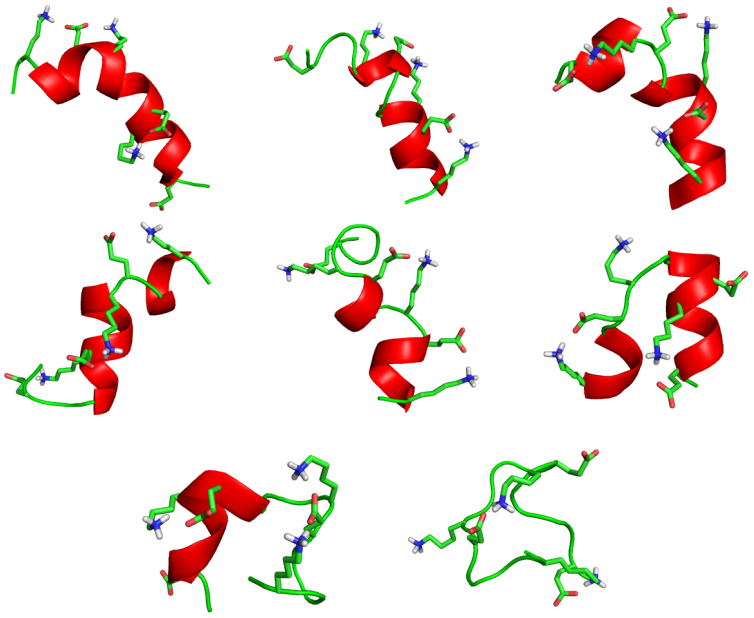


Figure 4.39: Eight unfolded replicas from a swarm directed simulated annealing simulation of AEK17, with a swarm potential of A = −25.0 kcal/mol.

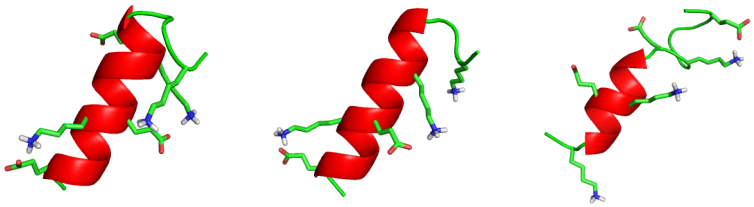


Figure 4.40: Three unfolded replicas from a swarm directed simulated annealing simulation of AEK17, with a swarm potential of A = −50.0 kcal/mol.

4.3.4.2 Reduction in Required Simulation Time

The use of two simulated annealing temperature schedules to fold Trp-cage allows us to comment on how the presence of a swarm potential reduces the simulation time required to fold the miniprotein to its native structure. With the shorter 25 ns schedule, two independent replicas displayed an average backbone RMSD of below 2.0 Å (Figure 4.14), relative to the reference NMR structure. The heavy atom RMSD of these two replicas was 2.78 and 3.02 Å (Figure 4.16). The Trp25 residue showed its correct orientation, ie pointing into the hydrophobic cage, in only one of these replicas (Figure 4.41). One further replica had a backbone RMSD of less than 2.3 Å and a correct Trp25 orientation. Two of the swarm directed replicas has an average backbone RMSD of below 2.0 Å (Figure 4.15), both of these had the correct Trp25 orientation (Figure 4.42), and their heavy atom RMSDs were 2.22 and 2.23 Å (Figure 4.17). 11 other replicas had backbone RMSDs of below 2.3 Å, and they all showed the correct Trp25 orientation.

Following the longer 40 ns simulated annealing schedule, four replicas folded to structures with an average backbone RMSD of below 2.0 Å (Figure 4.24), their heavy atom RMSDs were 2.83, 2.86, 2.95 and 3.12 Å (Figure 4.26). In three of these replicas, Trp25 adopted its correct orientation (Figure 4.43). With the swarm potential present, 17 of the replicas agreed with the backbone geometry of the reference structure to with 2.0 Å RMSD (Figure 4.25), with 16 of these having the correct Trp25 position, and their heavy atom RMSDs were in the range 2.1 – 3.0 Å.

The number of replicas that folded to within 2.0 Å backbone RMSD in the 25 ns swarm simulation and the 40 ns unbiased simulations, two and three, respectively, are comparable. The average heavy atom RMSD was also lower in these 25 ns

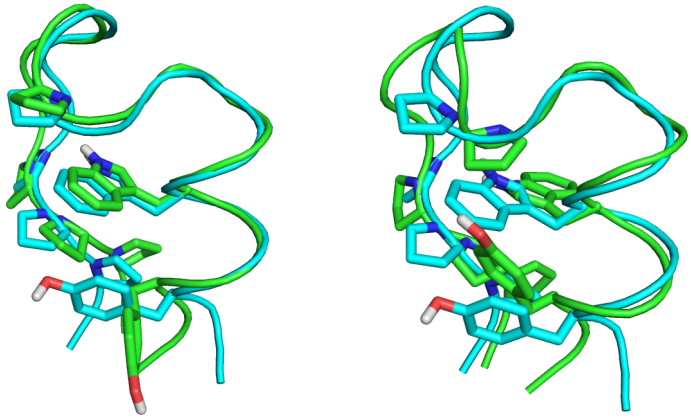


Figure 4.41: Two independent replicas of Trp-cage with the lowest backbone RMSD, obtained from 25 ns simulated annealing (green), compared to NMR reference structure (blue).

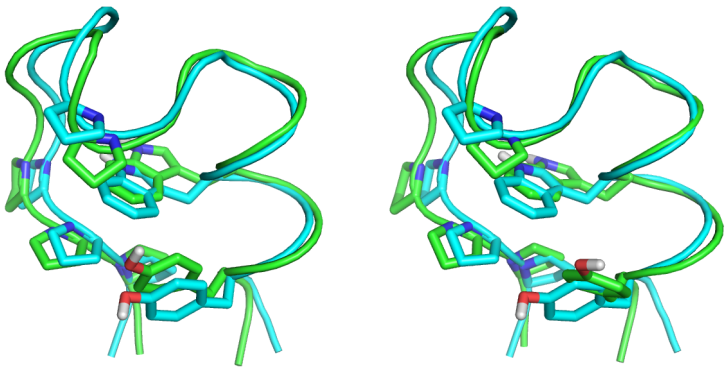


Figure 4.42: Two swarm directed replicas of Trp-cage with the lowest backbone RMSD, obtained from 25 ns simulated annealing (green), compared to NMR reference structure (blue).

swarm replicas. This was largely due to an incorrect Tyr22 position in two of the independent replicas (Figure 4.43), with the side chain of this residue rotated by approximately 180°.

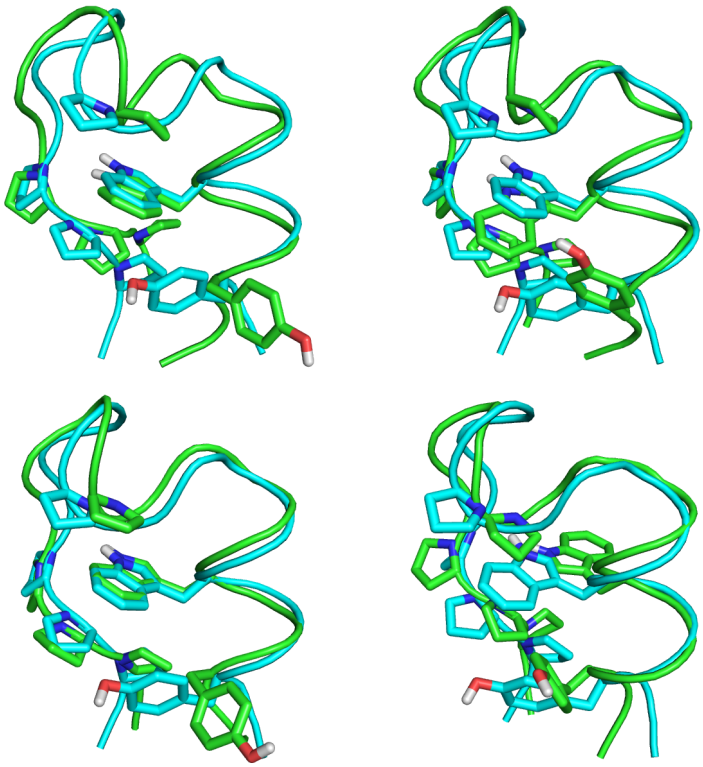


Figure 4.43: Four independent replicas of Trp-cage with the lowest backbone RMSD, obtained from 40 ns simulated annealing (green), compared to NMR reference structure (blue).

4.3.4.3 Frustration of Folded Replica Geometries by the Presence of Unfolded Replicas

Interestingly, while the final folded structures obtained from the 40 ns swarm simulation of Trp-cage showed less NOE restraint violations than the structures displayed in the independent replicas (Figure 4.31 and 4.30), the replica that showed the lowest number of violations was one of the independent replicas, with a fractional violation of 20%. The swarm directed replica with the least violations violated 23% of NOE restraints.

The reason for this seemingly worse performance by the swarm directed simu-

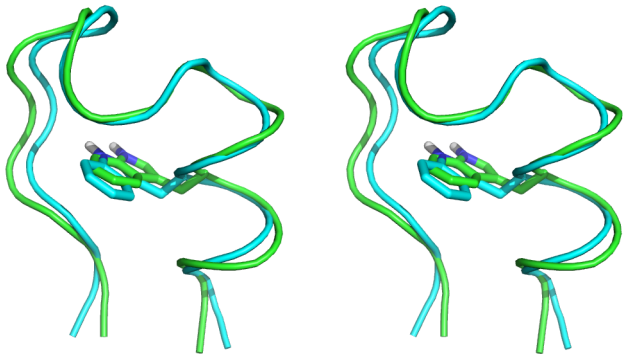


Figure 4.44: Stereoscopic view of displaced Trp25 residue in structure obtained from a 40 ns swarm directed simulation (green), compared to NMR reference (blue).

lation can be seen by examining the final folded structures. Of the 168 NOE restraints that were used to generate the NMR-derived structure of Trp-cage, 49 involved protons of the Trp25 residue.[110] In the folded structures obtained in the swarm simulation, this residue is subtly displaced, sitting slightly further back out of the hydrophobic cage structure (Figure 4.44). As a result, a number of restraints are violated. It seems that the presence of the four incorrectly folded simulation replicas is affecting the structures of the folded replicas.

By slowly tapering the swarm potential to zero over 2 ns, the negative effect of the swarm potential was removed, and the number of NOE restraint violations observed in the swarm directed replicas fell. The replica with the lowest number of violations in any of the independent replicas violated 17% of restraints (Figure 4.32), compared to 14% in the swarm replica with the least violations (Figure 4.33).

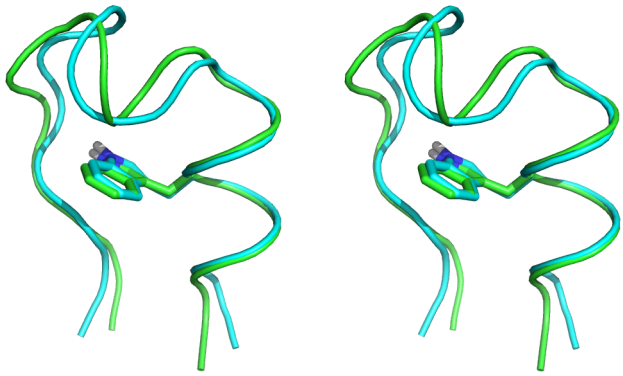


Figure 4.45: Steroscopic view of corrected Trp25 residue in structure following relaxation of swarm potential (green), compared to NMR reference (blue).

4.3.4.4 Comparison of Performance with Previous Studies

The performance, in terms of computational cost, of the swarm-based simulated annealing approach outlined here can be assessed by comparing to previously reported molecular dynamics studies. Chowdhury et al[115] performed 77 independent 100 ns folding simulations of Trp-cage, at 300 K, in implicit solvent. They reported that of these, only five folded to conformations with less than 2.0 Å backbone RMSD compared to the native state. Snow et al[116] used a distributed computed approach to perform over 1000 simulations of lengths longer than 30 ns and found that less than 1% folded to below 2.6 Å RMSD of the native C_α position. Our 20 replica swarm-directed simulation required a total simulation time of 800 ns (900 ns including final relaxation step) giving required simulation time per folded of 50 ns (56 ns). In comparison, Snow et al[116] required 1–4 μs of simulation time per folded replica.

Other advanced sampling approaches have also been applied to Trp-cage.[117–120] Kannan and Zacharias[120] performed a 40 ns temperature replica exchange simulation using 16 replicas at temperature between 300 – 460 K. They found that

the 300 K replica mainly sampled conformations with 2.0 Å RMSD of the native heavy atom structures, a comparable result to that achieved in this work. Hamiltonian replica exchange methods have also been applied to Trp-cage which only required five[120] or six[119] replicas. Replica exchange approaches have the additional benefit of sampling correct statistical ensembles, while the swarm approach does not. One drawback to these methods, however, is their poor scaling with system size, as efficient exchange between neighbouring replicas requires sufficient overlap in their energy distributions, as the method outlined here does not include any exchange steps, it is not bound by this requirement. This could potentially lead better performance when increasing the number of degrees of freedom in the system, for instance when including explicit solvent.

Chapter 5

Closing Remarks

5.1 Structure and Aggregation Inhibition of β-Amyloid Fibrils

From molecular dynamics simulations of Aβ in complex with peptides LPFFD and LHFFD, we found that the simulations could indeed distinguish the active peptide inhibitor LPFFD from the closely-related inactive one, LHFFD. The inactive LHFFD bound less strongly to the fibril than the active peptide, LPFFD, and formed distinct interactions. Whereas LPFFD interacted significantly with the fibril throughout the simulation, LHFFD, depending on initial location, detached from the fibril, to rebind at a later time point. Clustered bound peptide poses suggested that the ability to form inter-monomer hydrogen bonds that could disrupt the β-sheet formation, and in particular a key groove-bound orientation of LPFFD, could account for its activity. These results illustrate the potential of MD simulations as an aid to compound selection and design procedures, providing an alternate way to scoring functions in estimating the stability of interactions with the target.

To understand the energetic and mechanical features of the fibril receptor, we have also performed long explicit solvent MD simulations of unliganded amyloid fibril in three putative protonation states. Over these 100 ns simulations, the least deviation from the initial solid state NMR structures are exhibited by the fibrils in which the Glu11 and Glu22 side-chains are protonated. Subsequent free energy calculations on these trajectories suggest that the weakest fibril interface lies in the lateral rather than transverse direction. We also predict that there is little dependence on whether the transverse interface is at the edge or middle of the fibril, in agreement with reported steered molecular dynamics calculations. Alanine scanning provides residue-level detail into the factors contributing to

stability in the fibril. The contributions of nonpolar residues broadly agree with experimental alanine scanning results. However, the important of interactions between a number of polar residues appear to be overestimated. We suggest that shortcomings in the solvation model and in sampling of side-chains contributions may play a role in this observed discrepancy.

5.2 Swarm-based Molecular Dynamics

The conformational behaviour of biomolecules is typified by rough potential energy surfaces. In an effort to explore the ability of SWARM-MD to detect the low energy conformations of biomolecules, we have implemented the method in the Amber molecular dynamics suite, using non-periodic and periodic formulations of the influence of the swarm members on the simulated replicas. We apply the approach to three model systems, polyalanine, AEK17 and Trp-cage. In the absence of any swarm forces, a number of the independent replicas of polyalanine become trapped in non-helical conformations, with only five replicas folding correctly. Using SWARM-MD with a non-periodic dihedral angle definition, as described by Huber and van Gunsteren, did not improve the performance of the simulated annealing protocol. However, using the periodic dihedral angle definition increased polyalanine helix formation. For predicting the helicity of AEK17, only two independent replicas simulation folded to the native structure during the simulation. SWARM-MD calculations using a periodic dihedral formulations in general showed a marked improvement in helical formation compared to the independent replica simulations, achieving 12/20 or 17/20 fully folded replicas, depending on swarm potential strength. These potentials were again found to significantly outperform non-periodic potentials. Finally, improved numbers of

Trp-cage replicas folded correctly when the SWARM-MD potential was applied, either with a short or in particular longer simulated annealing protocol. However, it became evident that the presence of the four incorrectly folded simulation replicas affected the structures of the folded replicas. By slowly removing the swarm potential over 2 ns, the negative effect of the swarm potential was alleviated, and better agreement with the NMR structure of Trp-cage was obtained.

We note that other methods have been introduced to enhance sampling of phase space by molecular simulation methods, such as metadynamics, locally enhanced sampling and replica-exchange schemes. Several of these techniques have been applied to folding of Trp-cage. Most recently, a temperature REMD simulation was performed on Trp-cage in generalized Born solvent using sixteen 40 ns replicates spanning 300 – 460 K; the 300 K trajectory latterly sampled mainly folded Trp-cage (within about 2 Å heavy atom RMSD of the native structure). These calculation conditions and the method's performance are comparable to that of the SWARM-MD simulation of Trp-cage presented here. Hamiltonian-based replica exchange is also possible, and these methods have proved particularly powerful, for example obtaining folded Trp-cage structures using five or six replicas of sub-100 ns trajectories. An additional advantage of temperature-based and Hamiltonian-based replica exchange schemes is the generation of correct 300 K ensembles, providing information on folding pathways and intermediates. However, neither of these methods scale favourably with system size due to the constraint of the exchange condition: efficient exchange between neighbouring replicas require sufficient energy overlap between replicas. Therefore the number of required replicas grows rapidly with the size of the simulation system and correspondingly longer simulation times are required to allow efficient sampling of temperature space. SWARM-MD does not incorporate an exchange move and therefore does

not suffer from this exacting requirement. This difference may be of greater importance when extending the method to study explicitly solvated systems, where system size is greatly increased by the degrees of freedom of water molecules.

In conclusion, SWARM-MD, in its periodic formulation, is able to improve the ready detection of native structures of small – medium sized biomolecules. Given SWARM-MD algorithm's suitability for distribution of replicas over parallel architectures, the method appears has potential to predict the conformations of larger systems, particularly systems involving the incorporation of explicitly modeled bulk solvent. However, for future applications to predictive folding of larger polypeptide structures and interaction of flexible protein receptors with ligands, it will be useful to explore optimization schemes which anneal the influence of the swarm, to prevent unduly biasing outlier replicas.

References

[1] N. J. Bruce, D. Chen, S. G. Dastidar, G. E. Marks, C. H. Schein & R. A. Bryce (2010). Molecular dynamics simulations of Aβ fibril interactions with β-sheet breaker peptides. *Peptides*, **31**, 2100–2108.

[2] N. J. Bruce & R. A. Bryce (2010). Ab initio protein folding using a co-operative swarm of molecular dynamics trajectories. *Journal of Chemical Theory and Computation*, **6**, 1925–1930.

[3] J. E. Jones (1924). On the determination of molecular fields. II. from the equation of state of a gas. *Proceedings of the Royal Society of London. Series A*, **106**, 463–477.

[4] F. Jensen, *Introduction to Computational Chemisty*, chap. Force Field Methods, pp. 22–79. John Wiley & Sons, Ltd, 2007, second edn.

[5] F. London (1937). The general theory of molecular forces. *Trans. Faraday Soc.*, **33**, 8b–26.

[6] R. A. Buckingham (1938). The classical equation of state of gaseous helium, neon and argon. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, **168**, 264–283.

[7] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey & M. L.

Klein (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, **79**, 926–935.

[8] H. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren & J. Hermans, *Intermolecular Forces*, chap. Interaction Models for Water in Relation to Protein Hydration, pp. 331–342. Dordrecht, Reidel, 1983.

[9] H. J. C. Berendsen, J. R. Grigera & T. P. Straatsma (1987). The missing term in effective pair potentials. *Journal of Physical Chemistry*, **91**, 6269–6271.

[10] M. E. Davis & J. A. McCammon (1989). Solving the finite difference linearized Poisson-Boltzmann equation: A comparison of relaxation and conjugate gradient methods. *Journal of Computational Chemistry*, **10**, 386–391.

[11] A. Nicholls & B. Honig (1991). A rapid finite difference algorithm, utilizing successive over-relaxation to solve the poissonboltzmann equation. *Journal of Computational Chemistry*, **12**, 435–445.

[12] A. R. Leach, *Molecular Modelling: Principles and Applications*, chap. Three Challenges in Molecular Modelling, pp. 481–542. Longman, 1996.

[13] A. Onufriev, D. A. Case & D. Bashford (2002). Effective born radii in the generalized Born approximation: The importance of being perfect. *Journal of Computational Chemistry*, **23**, 1297–1304.

[14] V. Tsui & D. A. Case (2001). Theory and applications of the generalized Born solvation model in macromolecular simulations. *Biopolymers*, **56**, 275–291.

[15] V. Tsui & D. A. Case (2000). Molecular dynamics simulations of nucleic acids with a generalized born solvation model. *Journal of the American Chemical Society*, **122**, 2489–2498.

[16] A. Onufriev, D. Bashford & D. A. Case (2000). Modification of the generalized born model suitable for macromolecules. *Journal of Physical Chemistry B*, **104**, 3712–3720.

[17] G. D. Hawkins, C. J. Cramer & D. G. Truhlar (1995). Pairwise solute descreening of solute charges from a dielectric medium. *Chemical Physics Letters*, **246**, 122–129.

[18] G. D. Hawkins, C. J. Cramer & D. G. Truhlar (1996). Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *Journal of Physical Chemistry*, **100**, 19824–19839.

[19] F. Floris & J. Tomasi (1989). Evaluation of the dispersion contribution to the solvation energy. a simple computational model in the continuum approximation. *Journal of Computational Chemistry*, **10**, 616–627.

[20] F. M. Floris, J. Tomasi & J. L. P. Ahuir (1991). Dispersion and repulsion contributions to the solvation energy: Refinements to a simple computational model in the continuum approximation. *Journal of Computational Chemistry*, **12**, 784–791.

[21] L. Verlet (1967). Computer "experiments" on classical fluids. I. thermodynamical properties of Lennard-Jones molecules. *Physical Review*, **159**, 98.

[22] L. Verlet (1968). Computer "experiments"on classical fluids. II. equilibrium correlation functions. *Physical Review*, **165**, 201–214.

[23] M. P. Allen & D. J. Tildesley, *Computer Simulation of Liquids*, chap. Some tricks of the trade, pp. 140–182. Oxford University Press, 1987.

[24] T. Darden, D. M. York & L. Pedersen (1993). Particle mesh Ewald: An N.log(N) method for Ewald sums in large systems. *Journal Chemical Physics*, **98**, 10089–10092.

[25] H. C. Andersen (1980). Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of Chemical Physics*, **72**, 2384–2393.

[26] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola & J. R. Haak (1984). Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, **81**, 3684–3690.

[27] R. J. Loncharich, B. R. Brooks & R. W. Pastor (1992). Langevin dynamics of peptides: The frictional dependence of isomerization rates of N-acetylalanyl-N'-methylamide. *Biopolymers*, **32**, 523–535.

[28] D. Hamelberg, J. Mongan & J. A. McCammon (2004). Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *The Journal of Chemical Physics*, **120**, 11919–11929.

[29] A. Laio & M. Parrinello (2002). Escaping free-energy minima. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 12562–12566.

[30] C. Simmerling, T. Fox & P. A. Kollman (1998). Use of locally enhanced sampling in free energy calculations: Testing and application to the $\alpha \rightarrow \beta$ anomerization of glucose. *Journal of the American Chemical Society*, **120**, 5771–5782.

[31] Y. Sugita & Y. Okamoto (1999). Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters*, **314**, 141–151.

[32] T. Huber, A. E. Torda & W. F. Gunsteren (1994). Local elevation: A method for improving the searching properties of molecular dynamics simulation. *Journal of Computer-Aided Molecular Design*, **8**, 695–708.

[33] F. L. Gervasio, A. Laio & M. Parrinello (2005). Flexible docking in solution using metadynamics. *Journal of the American Chemical Society*, **127**, 2600–2607.

[34] D. A. Pearlman, D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham, S. DeBolt, D. Ferguson, G. Seibel & P. Kollman (1995). AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Computer Physics Communications*, **91**, 1–41.

[35] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller & E. Teller (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, **21**, 1087–1092.

[36] F. L. Mitchell, F. Frank, G. E. Marks, M. Suzuki, K. T. Douglas & R. A. Bryce (2009). Molecular dynamics study of chemically engineered green fluorescent protein mutants: Comparison of intramolecular fluorescence resonance energy transfer rate. *Proteins: Structure, Function, and Bioinformatics*, **75**, 28–39.

[37] J. Zhou, B. Liu, G. Geng & J. H. Wu (2010). Study of the impact of the t877a mutation on ligand-induced helix-12 positioning of the androgen

receptor resulted in design and synthesis of novel antiandrogens. *Proteins: Structure, Function, and Bioinformatics*, **78**, 623–637.

[38] H. Fukunishi, O. Watanabe & S. Takada (2002). On the hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. *The Journal of Chemical Physics*, **116**, 9058–9067.

[39] C. J. Woods, J. W. Essex & M. A. King (2003). The development of replica-exchange-based free-energy methods. *Journal of Physical Chemistry B*, **107**, 13703–13710.

[40] X. Cheng, G. Cui, V. Hornak & C. Simmerling (2005). Modified replica exchange simulation methods for local structure refinement. *Journal of Physical Chemistry B*, **109**, 8220–8230.

[41] T. Huber & W. F. van Gunsteren (1998). SWARM-MD: Searching conformational space by cooperative molecular dynamics. *Journal of Physical Chemistry A*, **102**, 5937–5943.

[42] R. W. Zwanzig (1954). High-temperature equation of state by a perturbation method. I. nonpolar gases. *Journal of Chemical Physics*, **22**, 1420–1426.

[43] P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case & T. E. Cheatham (2000). Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Accounts of Chemical Research*, **33**, 889–897.

[44] C. J. Cramer, *Essentials of Computational Chemistry: Theories and Models*, chap. Thermodynamic Properties, pp. 319–345. John Wiley & Sons, Ltd., 2002.

[45] K. Irikura, B. Tidor, B. Brooks & M. Karplus (1985). Transition from B to Z DNA: contribution of internal fluctuations to the configurational entropy difference. *Science*, **229**, 571–572.

[46] B. R. Brooks, D. Janeẑiŏ & M. Karplus (1995). Harmonic analysis of large systems. I. Methodology. *Journal of Computational Chemistry*, **16**, 1522–1542.

[47] R. Elber & M. Karplus (1987). Multiple conformational states of proteins: a molecular dynamics analysis of myoglobin. *Science*, **235**, 318–321.

[48] S. Hayward, A. Kitao & N. Gõ (1995). Harmonicity and anharmonicity in protein dynamics: A normal mode analysis and principal component analysis. *Proteins: Structure, Function, and Bioinformatics*, **23**, 177–186.

[49] B. J. Carrington & R. L. Mancera (2004). Comparative estimation of vibrational entropy changes in proteins through normal modes analysis. *Journal of Molecular Graphics and Modelling*, **23**, 167–174.

[50] M. Karplus & J. N. Kushick (1981). Method for estimating the configurational entropy of macromolecules. *Macromolecules*, **14**, 325–332.

[51] C.-E. Chang, W. Chen & M. K. Gilson (2005). Evaluating the accuracy of the quasiharmonic approximation. *Journal of Chemical Theory and Computation*, **1**, 1017–1028.

[52] K. Blennow, M. J. de Leon & H. Zetterberg (2006). Alzheimer's disease. *The Lancet*, **368**, 387 – 403.

[53] G. Rait, K. Walters, C. Bottomley, I. Petersen, S. Iliffe & I. Nazareth (2010). Survival of people with clinical diagnosis of dementia in primary care: cohort study. *British Medical Journal*, **341**.

[54] M. Ganguli, H. H. Dodge, C. Shen, R. S. Pandav & S. T. DeKosky (2005). Alzheimer disease and mortality: A 15-year epidemiological study. *Archives of Neurology*, **62**, 779–784.

[55] E. P. Helzner, N. Scarmeas, S. Cosentino, M. X. Tang, N. Schupf & Y. Stern (2008). Survival in Alzheimer disease. *Neurology*, **71**, 1489–1495.

[56] A. Ueki, H. Shinjo, H. Shimode, T. Nakajima & Y. Morita (2001). Factors associated with mortality in patients with early-onset Alzheimer's disease: a five-year longitudinal study. *International Journal of Geriatric Psychiatry*, **16**, 810–815.

[57] R. A. Stelzmann, H. Norman Schnitzlein & F. Reed Murtagh (1995). An English translation of Alzheimer's 1907 paper, " uber eine eigenartige erkankung der hirnrinde". *Clinical Anatomy*, **8**, 429–431.

[58] C. P. Ferri, M. Prince, C. Brayne, H. Brodaty, L. Fratiglioni, M. Ganguli, K. Hall, K. Hasegawa, H. Hendrie, Y. Huang, A. Jorm, C. Mathers, P. R. Menezes, E. Rimmer & M. Scazufca (2006). Global prevalence of dementia: a Delphi consensus study. *The Lancet*, **366**, 2112 – 2117.

[59] R. Wetzel (2006). Nucleation of huntingtin aggregation in cells. *Nature Chemical Biology*, **2**, 297 – 298.

[60] C. A. Davie (2008). A review of Parkinson's disease. *British Medical Bulletin*, **6**, 109–127.

[61] D. J. Selkoe (2003). Folding proteins in fatal ways. *Nature*, **426**, 900 – 904.

[62] C. L. Masters, G. Simms, N. A. Weinman, G. Multhaup, B. L. McDonald & K. Beyreuther (1985). Amyloid plaque core protein in Alzheimer disease

and Down syndrome. *Proceedings of the National Academy of Sciences*, **82**, 4245–4249.

[63] D. M. Holtzman, J. C. Morris & A. M. Goate (2011). Alzheimer's disease: The challenge of the second century. *Science Translational Medicine*, **3**, 77sr1.

[64] M. Morishima-Kawashima & Y. Ihara (2002). Alzheimer's disease: β-amyloid protein and tau. *Journal of Neuroscience Research*, **70**, 392–401.

[65] R. Kayed, E. Head, J. L. Thompson, T. M. McIntire, S. C. Milton, C. W. Cotman & C. G. Glabe (2003). Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis. *Science*, **300**, 486–489.

[66] R. M. Koffie, M. Meyer-Luehmann, T. Hashimoto, K. W. Adams, M. L. Mielke, M. Garcia-Alloza, K. D. Micheva, S. J. Smith, M. L. Kim, V. M. Lee, B. T. Hyman & T. L. Spires-Jones (2009). Oligomeric amyloid β associates with postsynaptic densities and correlates with excitatory synapse loss near senile plaques. *Proceedings of the National Academy of Sciences*, **106**, 4012–4017.

[67] S. Ho, L. Eckman, C. Long, K. Otvos Jr., L. Younkin, L. Suzuki, N. Younkin & S. Gravina (1995). Amyloid β protein (Aβ) in Alzheimer's disease brain. biochemical and immunocytochemical analysis with antibodies specific for forms ending at Aβ40 or Aβ42(43). *Journal of Biological Chemistry*, **270**, 7013–7016.

[68] O. M. A. El-Agnaf, D. S. Mahil, B. P. Patel & B. M. Austen (1993). Oligomerization and toxicity of β-amyloid–42 implicated in Alzheimer's dis-

ease. *Biochemical and Biophysical Research Communications*, **273**, 1003–1007.

[69] J. Hardy & D. J. Selkoe (2002). The amyloid hypothesis of Alzheimer's disease: Progress and problems on the road to therapeutics. *Science*, **297**, 353–356.

[70] C. Soto, E. M. Sigurdsson, L. Morelli, R. A. Kumar, E. M. Castano & B. Frangione (1998). β-sheet breaker peptides inhibit fibrillogenesis in a rat brain model of amyloidosis: Implications for Alzheimer's therapy. *Nature Medicine*, **4**, 822–826.

[71] D. Chen, Z. S. Martin, C. Soto & C. H. Schein (2009). Computational selection of inhibitors of Aβ aggregation and neuronal toxicity. *Bioorganic Medicinal Chemistry*, **17**, 5189–5197.

[72] A. T. Petkova, Y. Ishii, J. J. Balbach, O. N. Antzutkin, R. D. Leapman, F. Delaglio & R. Tycko (2002). A structural model for Alzheimer's β-amyloid fibrils based on experimental constraints from solid state NMR. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 16742–16747.

[73] A. T. Petkova, W. M. Yau & R. Tycko (2006). Experimental constraints on quaternary structure in Alzheimer's β-amyloid fibrils. *Biochemistry*, **45**, 498–512.

[74] D. S. Goodsell, G. M. Morris & A. J. Olson (1996). Automated docking of flexible ligands: Applications of AutoDock. *Journal of Molecular Recognition*, **9**, 1–5.

[75] G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K.

Belew & A. J. Olson (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry*, **19**, 1639–1662.

[76] D. A. Case, T. A. Darden, T. E. I. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, K. M. Merz, D. A. Pearlman, M. Crowley, R. C. Walker, W. Zhang, B. Wang, S. Hayik, A. Roitberg, G. Seabra, K. F. Wong, F. Paesani, X. Wu, S. Brozel, V. Tsui, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, P. Beroza, D. H. Mathews, C. Schafmeister, W. S. Ross & P. A. Kollman (2006), AMBER 9. University of California, San Francisco.

[77] Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. M. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. M. Wang & P. Kollman (2003). A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *Journal of Computational Chemistry*, **24**, 1999–2012.

[78] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey & M. L. Klein (1983). Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics*, **79**, 926–935.

[79] J.-P. Ryckaert, G. Ciccotti & H. J. C. Berendsen (1977). Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics*, **23**, 327–341.

[80] D. A. Case, T. A. Darden, T. E. I. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, M. Crowley, R. C. Walker, W. Zhang, K. M. Merz, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K. F. Wong,

F. Paesani, L. Vanicek, X. Wu, S. R. Brozel, T. Steinbrecher, H. Gohlke, , L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D. H. Mathews, M. G. Seetin, C. Sagui, V. Babin & P. A. Kollman (2008), AMBER 10. University of California, San Francisco.

[81] M. Feig, J. Karanicolas & C. L. I. Brooks (2004). MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. *Journal of Molecular Graphics and Modelling*, **22**, 377–395.

[82] J. B. Macqueen, Some methods of classification and analysis of multivariate observations. *In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, pp. 281–297, 1967.

[83] V. Tsui & D. A. Case (2000). Theory and applications of the generalized born solvation model in macromolecular simulations. *Biopolymers*, **56**, 275–291.

[84] C. Tan, L. Yang & R. Luo (2006). How well does Poisson-Boltzmann implicit solvent agree with explicit solvent? a quantitative analysis. *Journal of Physical Chemistry B*, **110**, 18680–18687.

[85] E. Gallicchio & R. M. Levy (2004). AGBNP: An analytic implicit solvent model suitable for molecular dynamics simulations and high-resolution modeling. *Journal of Computational Chemistry*, **25**, 479–499.

[86] C. Tan, Y.-H. Tan & R. Luo (2007). Implicit nonpolar solvent models. *Journal of Physical Chemistry B*, **111**, 12263–12274.

[87] W. Nadler & U. H. E. Hansmann (2008). Optimized explicit-solvent replica exchange molecular dynamics from scratch. *Journal of Physical Chemistry B*, **112**, 10386–10387.

[88] N. Huang, B. K. Shoichet & J. J. Irwin (2006). Benchmarking sets for molecular docking. *Journal of Medicinal Chemistry*, **49**, 6789–6801.

[89] A. Cavalli, G. Bottegoni, C. Raco, M. De Vivo & M. Recanatini (2004). A computational study of the binding of propidium to the peripheral anionic site of human acetylcholinesterase. *Journal of Medicinal Chemistry*, **47**, 3991–3999.

[90] Y. Pérez-Castillo, M. Froeyen, M. Cabrera-Pérez & A. Nowé (2011). Molecular dynamics and docking simulations as a proof of high flexibility in E.coli FabH and its relevance for accurate inhibitor modeling. *Journal of Computer-Aided Molecular Design*, pp. 1–23. 10.1007/s10822-011-9427-z.

[91] K. Lum, D. Chandler & J. D. Weeks (1999). Hydrophobicity at small and large length scales. *Journal of Physical Chemistry B*, **103**, 4570–4577.

[92] D. M. Huang & D. Chandler (2002). The hydrophobic effect and the influence of solute–solvent attractions. *Journal of Physical Chemistry B*, **106**, 2047–2053.

[93] C. Wu, Z. Wang, H. Lei, W. Zhang & Y. Duan (2007). Dual binding modes of Congo red to amyloid protofibril surface observed in molecular dynamics simulations. *Journal of the American Chemical Society*, **129**, 1225–1232.

[94] C. Wu, Z. Wang, H. Lei, Y. Duan, M. T. Bowers & J.-E. Shea (2008). The binding of thioflavin t and its neutral analog bta-1 to protofibrils of the alzheimer's disease A[beta]16-22 peptide probed by molecular dynamics simulations. *Journal of Molecular Biology*, **384**, 718–729.

[95] I. Massova & P. A. Kollman (1999). Computational alanine scanning to probe protein–protein interactions: A novel approach to evaluate binding free energies. *Journal of the American Chemical Society*, **121**, 8133–8143.

[96] E. P. Raman, T. Takeda, V. Barsegov & D. K. Klimov (2007). Mechanical unbinding of Aβ peptides from amyloid fibrils. *Journal of Molecular Biology*, **373**, 785–800.

[97] A. D. Williams, S. Shivaprasad & R. Wetzel (2006). Alanine scanning mutagenesis of Aβ(1-40) amyloid fibril stability. *Journal of Molecular Biology*, **357**, 1283–1294.

[98] K. Dill & H. Chan (1997). From Levinthal to pathways to funnels. *Nature Structural Biology*, **4**, 10–19.

[99] Y. Zhang & J. Skolnick (2005). The protein structure prediction problem could be solved using the current PDB library. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 1029–1034.

[100] M. Nanias, M. Chinchio, J. Pillardy, D. R. Ripoll & H. A. Scheraga (2003). Packing helices in proteins by global optimization of a potential energy function. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 1706–1710.

[101] G. Jones, P. Willett, R. C. Glen, A. R. Leach & R. Taylor (1997). Development and validation of a genetic algorithm for flexible docking. *Journal of Molecular Biology*, **267**, 727 – 748.

[102] G. Beni, From swarm intelligence to swarm robotics. *In* E. Sahin & W. Spears (eds.), *Swarm Robotics*, vol. 3342 of *Lecture Notes in Computer Science*, pp. 1–9, Springer Berlin / Heidelberg, 2005.

[103] O. Korb, T. Stützle & T. Exner, PLANTS: Application of ant colony optimization to structure-based drug design. *In* M. Dorigo, L. Gambardella, M. Birattari, A. Martinoli, R. Poli & T. Stützle (eds.), *Ant Colony Opti-*

mization and Swarm Intelligence, vol. 4150 of *Lecture Notes in Computer Science*, pp. 247–258, Springer Berlin / Heidelberg, 2006.

[104] Z. Xiang, C. S. Soto & B. Honig (2002). Evaluating conformational free energies: The colony energy and its application to the problem of loop prediction. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 7432–7437.

[105] J. Kennedy & R. Eberhart, Particle swarm optimization. *In Neural Networks, 1995. Proceedings., IEEE International Conference on*, vol. 4, pp. 1942–1948, 1995.

[106] W. Cedeño & D. K. Agrafiotis (2003). Using particle swarms for the development of QSAR models based on K-nearest neighbor and kernel regression. *Journal of Computer-Aided Molecular Design*, **17**, 255–263.

[107] H.-M. Chen, B.-F. Liu, H.-L. Huang, S.-F. Hwang & S.-Y. Ho (2007). SODOCK: Swarm optimization for highly flexible protein–ligand docking. *Journal of Computational Chemistry*, **28**, 612–623.

[108] N. I. Fisher, *Statistical analysis of circular data*, chap. Descriptive methods, pp. 15–37. Cambridge University Press, 1993.

[109] H. Bekker, H. J. C. Berendsen & W. F. van Gunsteren (1995). Force and virial of torsional-angle-dependent potentials. *Journal of Computational Chemistry*, **16**, 527–533.

[110] J. Neidigh, R. Fesinmeyer & N. Andersen (2002). Designing a 20-residue protein. *Nature Stuctural Biology*, **9**, 425–430.

[111] Y. Levy, J. Jortner & O. M. Becker (2001). Solvent effects on the energy

landscapes and folding kinetics of polyalanine. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 2188–2193.

[112] S. Marqusee & R. L. Baldwin (1987). Helix stabilization by glu−...lys+ salt bridges in short peptides of de novo design. *Proceedings of the National Academy of Sciences of the United States of America*, **84**, 8898–8902.

[113] W. Kabsch & C. Sander (1983). Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.

[114] W.-Z. Wang, T. Lin & Y.-C. Sun (2007). Examination of the folding of a short alanine-based helical peptide with salt bridges using molecular dynamics simulations. *Journal of Physical Chemistry B*, **111**, 3508–3514.

[115] S. Chowdhury, M. C. Lee & Y. Duan (2004). Characterizing the rate-limiting step of Trp-cage folding by all-atom molecular dynamics simulations. *Journal of Physical Chemistry B*, **108**, 13855–13865.

[116] C. D. Snow, B. Zagrovic & V. S. Pande (2002). The Trp cage: Folding kinetics and unfolded state topology via molecular dynamics simulations. *Journal of the American Chemical Society*, **124**, 14548–14549.

[117] J. W. Pitera & W. Swope (2003). Understanding folding and design: Replica-exchange simulations of "Trp-cage" miniproteins. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 7587–7592.

[118] S. Piana & A. Laio (2007). A bias-exchange approach to protein folding. *Journal of Physical Chemistry B*, **111**, 4553–4559.

[119] W. J. Son, S. Jang, Y. Pak & S. Shin (2007). Folding simulations with novel conformational search method. *Journal of Chemical Physics*, **126**.

[120] S. Kannan & M. Zacharias (2009). Folding of Trp-cage mini protein using temperature and biasing potential replica-exchange molecular dynamics simulations. *International Journal of Molecular Sciences*, **10**, 1121–1137.